

**2004**  
**Grade Eight Proficiency**  
**Assessment (GEPA)**

**TECHNICAL REPORT**  
**March Administration**



PTM#1505.95

Copyright © 2006 New Jersey

Department of Education

All rights reserved.





<b>CHAPTER 1: INTRODUCTION</b> . . . . .	<b>1</b>
1.1 Description of the Grade Eight Proficiency Assessment (GEPA) . . . . .	1
1.2 Purpose of the GEPA . . . . .	3
1.3 GEPA Organizational Support. . . . .	5
<b>CHAPTER 2: TEST DEVELOPMENT</b> . . . . .	<b>6</b>
2.1 Test Specifications. . . . .	6
Language Arts Literacy . . . . .	8
Mathematics . . . . .	8
Science . . . . .	9
2.2 Development of Test Items . . . . .	9
2.3 Item Review Process. . . . .	12
2.4 Operational Test Development . . . . .	21
2.5 Review and Approve Operational Test Forms . . . . .	22
2.6 Test Materials for Visually Impaired Students. . . . .	22
<b>CHAPTER 3: TEST ADMINISTRATION</b> . . . . .	<b>23</b>
3.1 Participation. . . . .	23
3.2 Test Security Procedures. . . . .	24
3.3 Test Administration Procedures. . . . .	26
3.4 Test Accommodations . . . . .	27
<b>CHAPTER 4: SCORING</b> . . . . .	<b>29</b>
4.1 Multiple-choice Items . . . . .	29
4.2 Open-ended Items . . . . .	29
Scorer Selection . . . . .	30
Rangefinding. . . . .	30
Development of Scoring Guides . . . . .	31
Team Leader Training and Qualifying . . . . .	31
Scorer Training and Qualifying . . . . .	32
Scoring Procedures and Paper Flow . . . . .	33
Scorer Monitoring . . . . .	34
Agreement Between Scorers for the Writing Tasks and Open-Ended Items. . . . .	34
4.3 Quality Control Procedures in Data Preparation . . . . .	36
Scanning and Scoring . . . . .	36
NJDOE Quality Control of Score Reporting . . . . .	37
<b>CHAPTER 5: STANDARD SETTING</b> . . . . .	<b>39</b>
5.1 Overview of the Process . . . . .	39

<b>5.2 Procedures</b> . . . . .	<b>40</b>
<i>Language Arts Literacy–Proficiency Level Descriptors</i> . . . . .	<b>40</b>
<i>Mathematics–Proficiency Level Descriptors</i> . . . . .	<b>41</b>
<i>Science–Proficiency Level Descriptors</i> . . . . .	<b>42</b>
<i>Judge Selection Process and Criteria</i> . . . . .	<b>43</b>
<i>Holistic/Paper Sorting Methodology</i> . . . . .	<b>43</b>
<i>Overview of the 8-Step Plan</i> . . . . .	<b>44</b>
<b>5.3 Results</b> . . . . .	<b>46</b>
<b>CHAPTER 6: SCALING AND EQUATING</b> . . . . .	<b>48</b>
<b>6.1 Scaling</b> . . . . .	<b>48</b>
<b>6.2 Equating</b> . . . . .	<b>48</b>
<i>Mathematics and Science Equating Design</i> . . . . .	<b>49</b>
<i>Language Arts Literacy Equating Design</i> . . . . .	<b>50</b>
<b>CHAPTER 7: TEST STATISTICS</b> . . . . .	<b>52</b>
<b>7.1 Reliability of the Test Scores</b> . . . . .	<b>52</b>
<b>CHAPTER 8: ITEM-LEVEL STATISTICS</b> . . . . .	<b>55</b>
<b>8.1 Classical Item Statistics</b> . . . . .	<b>55</b>
<b>8.2 Speededness</b> . . . . .	<b>58</b>
<b>8.3 Intercorrelations</b> . . . . .	<b>60</b>
<b>CHAPTER 9: TEST VALIDITY</b> . . . . .	<b>62</b>
<b>9.1 Content and Curricular Validity</b> . . . . .	<b>64</b>
<b>9.2 Construct Validity</b> . . . . .	<b>65</b>
<b>9.3 Criterion-Related Validity</b> . . . . .	<b>66</b>
<b>9.4 Consequential Validity Evidence</b> . . . . .	<b>67</b>
<b>CHAPTER 10: REPORTING</b> . . . . .	<b>71</b>
<b>10.1 Information on the Reports</b> . . . . .	<b>71</b>
<b>10.2 Types of Reports</b> . . . . .	<b>73</b>
<i>Cycle I Reports</i> . . . . .	<b>73</b>
<i>Cycle II Reports</i> . . . . .	<b>78</b>
<b>Appendix A: Scoring Rubrics and 3rd Reader Score Calculation Charts</b> . . . . .	<b>82</b>
<b>Appendix B: Cycle II Test Results</b> . . . . .	<b>93</b>
<b>Appendix C: Raw to Scale Scores Conversions</b> . . . . .	<b>111</b>
<b>Appendix D: Scale Scores with Frequencies</b> . . . . .	<b>121</b>
<b>Appendix E: Limited English Proficiency (LEP) and Special Education (SE)</b> . . . . .	<b>128</b>
<b>References</b> . . . . .	<b>130</b>

## Tables

<b>TABLE 1.1 Number of Items and Approximate Times . . . . .</b>	<b>1</b>
<b>TABLE 1.2 Total Student Group Testing in 2004 . . . . .</b>	<b>2</b>
<b>TABLE 2.1 Total Points Possible for the Language Arts Literacy Component of the GEPA . . . . .</b>	<b>10</b>
<b>TABLE 2.2 Total Points Possible for the Mathematics Component of the GEPA . . . . .</b>	<b>10</b>
<b>TABLE 2.3 Total Points Possible for the Science Component of the GEPA . . . . .</b>	<b>10</b>
<b>TABLE 2.4 Item Development Goals for the 2004 Field Test . . . . .</b>	<b>12</b>
<b>TABLE 2.5 District Factor Groups (DFG) Represented on the GEPA Content and Sensitivity Committees . . . . .</b>	<b>13</b>
<b>TABLE 2.6 GEPA 2004 Content and Sensitivity Committee Meetings . . . . .</b>	<b>14</b>
<b>TABLE 2.7 Number of Field Test Items Approved During Statistical Review-LANGUAGE ARTS LITERACY-READING . . . . .</b>	<b>17</b>
<b>TABLE 2.8 Number of Field Test Items Approved During Statistical Review-MATHEMATICS . . . . .</b>	<b>19</b>
<b>TABLE 2.9 Number of Field Test Items Approved During Statistical Review Classified with 2002/2004 Core Curriculum Content Standards-MATHEMATICS . . . . .</b>	<b>19</b>
<b>TABLE 2.10 Number of Field Test Items Approved During Statistical Review-SCIENCE . . . . .</b>	<b>20</b>
<b>TABLE 2.11 Number of Field Test Items Approved During Statistical Review Classified with 2002/2004 Core Curriculum Content Standards-SCIENCE . . . . .</b>	<b>20</b>
<b>TABLE 2.12 Operational Test Specifications . . . . .</b>	<b>21</b>
<b>TABLE 4.1 Number of Writing Prompt and Open-ended Items Scored . . . . .</b>	<b>29</b>
<b>TABLE 4.2 Consistency Between Raters Scoring GEPA Writing Tasks and Open-Ended Items - 2004. . . . .</b>	<b>35</b>
<b>TABLE 5.1 Proficiency-Level Cut Scores . . . . .</b>	<b>47</b>
<b>TABLE 5.2 Percentage of Students Achieving Each Performance Level. . . . .</b>	<b>47</b>
<b>TABLE 7.1 Reliability Estimates and Standard Errors of Measurement (SEMs) for GEPA Language Arts Literacy, Mathematics, and Science - 2004. . . . .</b>	<b>53</b>
<b>TABLE 7.2 Kuder-Richardson 20 Reliability Estimates and Standard Errors of Measurement (SEMs) for Dichotomously Scored Items Within GEPA Content Clusters - 2004. . . . .</b>	<b>54</b>

<b>TABLE 8.1 Item Difficulty and Discrimination Summary Statistics for Dichotomously Scored and Open-Ended Items by Test Section and Cluster - 2004</b> . . . . .	<b>55</b>
<b>TABLE 8.2 Frequency Distributions of Item Difficulty and Discrimination by Content Cluster - 2004 Language Arts Literacy</b> . . . . .	<b>56</b>
<b>TABLE 8.3 Frequency Distributions of Item Difficulty and Discrimination by Content Cluster - 2004 Mathematics</b> . . . . .	<b>57</b>
<b>TABLE 8.4 Frequency Distributions of Item Difficulty and Discrimination by Content Cluster - 2004 Science</b> . . . . .	<b>58</b>
<b>TABLE 8.5 Percentage of Students Omitting the Last Items of Each Test Part - 2004</b> . . . . .	<b>59</b>
<b>TABLE 8.6 Intercorrelations Among GEPA Major Content Clusters and Item Types - 2004</b> . . . . .	<b>60</b>
<b>TABLE 8.7 Intercorrelations Among GEPA Content Areas and Clusters - 2004</b> . . . . .	<b>61</b>

## Figures

<b>FIGURE 2.1: GEPA Test Development Process</b> . . . . .	<b>7</b>
<b>FIGURE 2.2: Item Approval Before Field Test</b> . . . . .	<b>13</b>
<b>FIGURE 2.3: Item Approval Before Operational Test</b> . . . . .	<b>16</b>
<b>FIGURE 9.1 LANGUAGE ARTS LITERACY - Longitudinal Graph by Economic Status</b> . . . . .	<b>68</b>
<b>FIGURE 9.2 MATHEMATICS - Longitudinal Graph by Economic Status</b> . . . . .	<b>68</b>
<b>FIGURE 9.3 SCIENCE - Longitudinal Graph by Economic Status</b> . . . . .	<b>69</b>
<b>FIGURE 10.1: ISR (Front)</b> . . . . .	<b>74</b>
<b>FIGURE 10.2: ISR (Back)</b> . . . . .	<b>75</b>
<b>FIGURE 10.3: Student Roster (Mathematics)</b> . . . . .	<b>77</b>
<b>FIGURE 10.4: Cluster Means Report</b> . . . . .	<b>79</b>
<b>FIGURE 10.5: Performance by Demographic Groups</b> . . . . .	<b>81</b>



## CHAPTER 1: INTRODUCTION

### 1.1 Description of the Grade Eight Proficiency Assessment (GEPA)

The New Jersey Grade Eight Proficiency Assessment (GEPA) for the 2004 administration consisted of three content area tests - Language Arts Literacy, Mathematics, and Science. The GEPA is designed to provide an indication of the progress students are making in mastering the knowledge and skills described in New Jersey's Core Curriculum Content Standards for these content areas.

The GEPA was administered between Monday, March 8 and Thursday, March 11, 2004, with make-up testing between Monday, March 15 and Thursday, March 18, 2004. Table 1.1 lists the number of test items and approximate testing time for the three content areas.

**TABLE 1.1**  
**Number of Items and Approximate Times**

<b>Content Areas</b>	<b>Items</b>	<b>Approximate Times and Days</b>
Science	45 multiple-choice 3 open-ended Embedded field-test items	Monday morning  1 hour, 57 minutes
Mathematics	30 multiple-choice 6 open-ended Embedded field-test items	Tuesday morning  2 hours, 27 minutes
Language Arts Literacy	20 multiple-choice 4 open-ended 2 writing tasks 1 revising/editing passage Field-test component	Wednesday morning Thursday morning  2 hours, 12 minutes (per day)

The GEPA Language Arts Literacy measures both reading and writing. The Reading component requires students to read passages and to respond to related items. The passages are selected from published books, newspapers, and magazines, as well as everyday text. The Reading component includes both multiple-choice and open-ended items. The open-ended items require students to write a few sentences or a few paragraphs to answer a question about the text. The Writing component asks students to write two essays. All the tasks in the Writing component require students to write a response.

The GEPA Mathematics measures students' abilities to solve problems using mathematical concepts. The components in this content area measure: Number Sense, Concepts, and Applications; Spatial Sense and Geometry; Data Analysis, Probability, Statistics, and Discrete Mathematics; and Patterns, Functions, and Algebra. Mathematics, like the Reading component of Language Arts Literacy, contains both multiple-choice and open-ended items. The open-ended items require students to solve a problem as well as explain their solution.

The GEPA Science measures students’ knowledge and skills in Life Science, Physical Science, and Earth Science. The Science content area contains both multiple-choice and open-ended items. The open-ended items require students to respond to a question as well as explain the answer.

Rubrics for scoring the GEPA open-ended items and writing prompts are included in Appendix A of this Technical Report.

Table 1.2 presents the statewide test results for the 2004 administration of the GEPA. This table shows the number and percentages of students in each of the Proficiency Levels – Partially Proficient, Proficient, and Advanced Proficient. The first column in Table 1.2 shows the total 110,270 enrolled students including 88,480 general education students, 18,010 special education students, and 3,982 limited English students. “General Education” excludes students coded as special education (SE) or limited English proficient (LEP) on their answer folders. “Special Education” includes students coded as SE. “Limited English Proficient” includes students coded as LEP. “Total Students” refers to all students tested.

Following the Number Enrolled column are the columns for Number Not Present and Number of Voids. Number enrolled represents total number of answer folders returned. Number not present indicates the number of answer folders returned that were totally blank excluding answer folders coded as APA/IEP Exempt. A student’s answer folder can be voided at the time of testing due to illness, cheating or disruptive behavior, or some other reason. If a student’s answer folder is voided, no total test score for that student is reported for the content area. A Void code is printed in place of the total test score on the student’s individual reports.

**TABLE 1.2**  
**Total Student Group Testing in 2004**

TESTS	NUMBER ENROLLED	NUMBER NOT PRESENT	NUMBER OF VOIDS	NUMBER OF VALID SCALE SCORES	PROFICIENCY LEVELS						MEAN SCALE SCORE
					PARTIALLY PROFICIENT (100 – 199)		PROFICIENT (200 – 249)		ADVANCED PROFICIENT (250 – 300)		
					NO.	%	NO.	%	NO.	%	
<b>LANGUAGE ARTS LITERACY</b>											
General Education	88,480	277	424	87,779	15,350	17.5	66,540	75.8	5,889	6.7	219.4
Special Education	18,010	230	837	16,943	12,257	72.3	4,648	27.4	38	0.2	181.4
Limited English Proficient	3,982	39	50	3,893	3,211	82.5	675	17.3	7	0.2	171.0
<b>Total Students</b>	<b>110,270</b>	<b>544</b>	<b>1,299</b>	<b>108,427<sup>a</sup></b>	<b>30,641</b>	<b>28.3</b>	<b>71,852</b>	<b>66.3</b>	<b>5,934</b>	<b>5.5</b>	<b>211.9</b>
<b>MATHEMATICS</b>											
General Education	88,480	353	82	88,045	25,276	28.7	41,525	47.2	21,244	24.1	220.2
Special Education	18,010	296	530	17,184	13,607	79.2	3,170	18.4	407	2.4	180.1
Limited English Proficient	3,982	40	14	3,928	3,003	76.5	745	19.0	180	4.6	182.4
<b>Total Students</b>	<b>110,270</b>	<b>687</b>	<b>618</b>	<b>108,965<sup>b</sup></b>	<b>41,717</b>	<b>38.3</b>	<b>45,420</b>	<b>41.7</b>	<b>21,828</b>	<b>20.0</b>	<b>212.6</b>
<b>SCIENCE</b>											
General Education	88,480	446	83	87,951	15,172	17.3	50,664	57.6	22,115	25.1	228.1
Special Education	18,010	398	448	17,164	9,595	55.9	6,753	39.3	816	4.8	199.0
Limited English Proficient	3,982	47	19	3,916	2,895	73.9	968	24.7	53	1.4	188.2
<b>Total Students</b>	<b>110,270</b>	<b>886</b>	<b>543</b>	<b>108,841<sup>c</sup></b>	<b>27,502</b>	<b>25.3</b>	<b>58,356</b>	<b>53.6</b>	<b>22,983</b>	<b>21.1</b>	<b>222.2</b>

<sup>a</sup> The number of Valid Scale Scores includes 188 students who are both Special Education and Limited English Proficient.

<sup>b</sup> The number of Valid Scale Scores includes 192 students who are both Special Education and Limited English Proficient.

<sup>c</sup> The number of Valid Scale Scores includes 190 students who are both Special Education and Limited English Proficient.

During the scoring process, a void code is given if a student's answer folder showed less than 20 percent of the items attempted on the Mathematics or Science content area tests. During the 2004 administration, 445 Mathematics and 102 Science tests were voided due to the attempted criteria.

For Language Arts Literacy, if a student attempted less than 20 percent of the items on one or two testing days but attempted 20 percent or more on the other testing day, a Void code appeared instead of a total test score on the student's reports. However, cluster scores are provided for parts of the Language Arts Literacy which are attempted. During the 2004 administration, 256 Language Arts Literacy tests were voided due to the attempted criteria for Day 1 and 316 Language Arts Literacy tests were voided due to the attempted criteria for Day 2.

Table 1.2 shows that a total of 108,427 students had valid scale scores on Language Arts Literacy, 108,965 students had valid scale scores on Mathematics, and 108,841 students had valid scale scores on Science. The number of valid scale scores is the number enrolled excluding not-present and voids.

Performance data shown in the Proficiency Levels columns include students who received valid scale scores. The number of students who scored in each proficiency level excludes students coded as APA/IEP Exempt. Because each content area is independent, students may receive a scale score in one content area, but not in others.

The total GEPA Language Arts Literacy, Mathematics, and Science scores are reported as scale scores with a range of 100 to 300. Scale scores of 100 and 300 are a theoretical floor and ceiling, which may not actually be observed. The score ranges are as follows:

<i>Advanced Proficient</i>	<i>250–300</i>
<i>Proficient</i>	<i>200–249</i>
<i>Partially Proficient</i>	<i>100–199</i>

A series of tables summarizing the test results for the State (general education students, special education students, limited English proficient students, and total students), District Factor Groups, Special Needs Districts, and All Other (Non Special Needs) Districts appears in Appendix B. See <http://www.state.nj.us/njded/finance/> for information about District Factor Groups and Special Needs Districts (Abbott Districts).

**NOTE: Percentages shown in tables through this Technical Report may not total 100 due to rounding.**

## 1.2 Purpose of the GEPA

The GEPA serves as a primary indicator for identifying those students who may need instructional intervention in the three content areas of Language Arts Literacy, Mathematics, and Science. The test also serves as an indicator for determining which local education programs may require revisions to ensure that instructional programs are aligned with the Core Curriculum Content Standards. The GEPA is designed to evaluate the progress students are making in mastering the knowledge and skills required by the end of eighth grade. Also, the GEPA provides an indication of students' progress in the skills required to pass the High School Proficiency Assessment.

Three proficiency levels have been determined for each of the content areas of the GEPA: Partially Proficient, Proficient, and Advanced Proficient. Students scoring in the lowest level, Partially Proficient, are considered below the state minimum level of proficiency. These students may need instructional intervention. Instructional decisions for all students are determined only after additional information is considered, e.g., classroom tests, teacher observations.

In 1996, the State Board of Education adopted Core Curriculum Content Standards to describe what all students should know and be able to do at the end of fourth grade, eighth grade, and upon completion of a New Jersey public school education. The Core Curriculum Standards delineate New Jersey's expectations for student learning. All New Jersey school districts are required to organize instruction and design curricula so that virtually all students achieve these content standards. The Core Curriculum Content Standards defined the development of three statewide assessments: the Elementary School Proficiency Assessment Program, which was administered from 1997-2002; the GEPA, which replaced the Early Warning Test in 1998; and the High School Proficiency Assessment, which replaced the High School Proficiency Test as the state's graduation requirement for all students who entered the eleventh grade in the fall of 2001.

Previously, in 1988, the New Jersey Legislature passed a law which established the Early Warning Test. The Legislature moved the High School Proficiency Test from the ninth grade to the eleventh grade. The Grade 11 High School Proficiency Test assessed essential Reading, Mathematics, and Writing skills. It served as a graduation requirement for all public school students in New Jersey who entered ninth grade on or after September 1, 1991, and prior to fall of 2001.

The Early Warning Test was similar to the High School Proficiency Test in eleventh grade because it also measured basic skills in reading, mathematics, and writing. The Early Warning Test was administered to all eighth-grade students each spring to determine whether they were making satisfactory progress in mastering the skills they would need to pass the High School Proficiency Test in the eleventh grade. The Early Warning Test was first administered as an operational test in March 1994.

Following the adoption of the Core Curriculum Standards in 1996, the development of the GEPA was defined. The GEPA was initially administered as field tests in Language Arts Literacy and Mathematics. In March 1999, the GEPA was administered for the first time as an operational assessment. Additional field tests in Language Arts Literacy, Mathematics, and Science were also administered and the GEPA Speaking assessment was pilot tested. In March 2000, Science was included in GEPA as an operational test for the first time.

Because the State Board required that the Core Curriculum Content Standards be reviewed and revised every five years, a review process began in May 2001 involving teachers, school administrators, students, parents, and representatives from business, higher education, and the community.

The language arts literacy, mathematics, and science standards were adopted by the State Board of Education in July 2002. In April 2004, the language arts literacy standards were revised to comply with the requirements of the No Child Left Behind Act of 2001 (NCLB) and readopted by the Board.

The GEPA administration in 2004 included field test items which were aligned with the new Core Curriculum Content Standards for language arts literacy, mathematics, and science. The GEPA test development procedures are detailed in Chapter 2 of this manual.

### 1.3 GEPA Organizational Support

**New Jersey Department of Education (NJDOE)** The GEPA is administered by the Office of Evaluation and Assessment within the Department of Education. The staff of the Office of Evaluation and Assessment directs the implementation of the statewide assessment programs. In addition to planning, scheduling, and directing all GEPA activities, the staff is extensively involved in numerous test review, security, and quality control procedures.

**Pearson Educational Measurement (PEM—previously NCS Pearson)** In 1998, the contract for developing and administering the GEPA was awarded to NCS Pearson which became Pearson Educational Measurement in 2003. Pearson Educational Measurement is the primary contractor working in partnership with Measurement Incorporated (MI) and Assessment and Evaluation Services (AES). Major Pearson Educational Measurement activities include the following:

- Supporting and monitoring the test development cycle and subcontractor efforts toward content development
- Printing test books and ancillary materials required for the GEPA
- Distributing assessment materials in a secure manner and in appropriate amounts based on the district quantity survey results
- Supporting the regional workshops that inform district test coordinators about the GEPA program
- Receiving, scanning, editing, and scoring the answer documents using clearly defined quality control procedures
- Packaging and transporting open-ended responses to be hand-scored
- Providing accurate reports of test results to New Jersey pupils, parents/guardians, schools, districts, and the state

**Measurement Incorporated (MI)** MI provides item development and scores all open-ended responses for the GEPA program. Items developed include multiple-choice and constructed-response items for Language Arts Literacy, Mathematics, and Science; and writing prompts for Language Arts Literacy. MI scoring directors, NJDOE Office of Evaluation and Assessment content specialists, and New Jersey teachers use rangefinding procedures to prepare for scoring the GEPA open-ended items.

**Assessment and Evaluation Services (AES)** AES is responsible for GEPA technical activities such as specifying the item selection for the operational tests, equating the test forms, and developing the scale score conversion tables.

## CHAPTER 2: TEST DEVELOPMENT

The New Jersey Department of Education has developed a comprehensive set of assessments that measure student achievement of the Core Curriculum Content Standards. The validity of the GEPA is therefore based on the alignment of the GEPA, the Core Curriculum Content Standards, and the knowledge and skills expected of eighth-grade students.

The *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999, p. 11-12) notes the following possible sources of validity evidence:

- Evidence based on test content
- Evidence based on response processes
- Evidence based on internal structure
- Evidence based on relations to other variables
- Evidence based on consequences of testing

For an assessment like GEPA, content validity is the most relevant and important source of evidence. This chapter presents validity evidence based on test content. A description of the test specification development is followed by the procedures for test item development. Details about item writing as well as task, prompt, and passage selection are included. The last section delineates the review work of the New Jersey Assessment Content Committees. Additionally, an external committee assisted the New Jersey Department of Education by reviewing the assessments to determine how well they measure the knowledge and skills stated in the standards, and by comparing the New Jersey standards with those in other states and countries.

Chapter 9 of this *Technical Report* addresses validity evidence based on internal structure of the test. Item statistics and intercorrelations provide validity evidence related to internal structure.

### 2.1 Test Specifications

The GEPA content areas of Language Arts Literacy, Mathematics, and Science were designed from their inception in 1997 to align with the original Core Curriculum Content Standards adopted by the New Jersey State Board of Education in 1996. The State Board required that the Core Curriculum Content Standards be reviewed every five years. New standards for the three content areas were adopted by the Board in July 2002. To comply with requirements of the federal No Child Left Behind Act of 2001 (NCLB), the Language Arts Literacy standards were also revised in April 2004.

The Core Curriculum Content Standards were developed by teachers and other educational professionals from New Jersey. The Core Curriculum Content Standards outline what students should know and be able to do at a certain grade level. The questions on the GEPA can contain items/concepts included in the grade eight standards as well as for those standards listed in the prior grade standards.

The GEPA was first administered as an operational assessment in 1999. Prior to that time, the GEPA Language Arts Literacy and Mathematics was administered to all eighth-grade students as field tests and “due-notice” administrations. Science was initially field tested in 1999. The purpose of due-notice administrations was to help school districts identify potential gaps between their curriculum and the test objectives, and to allow schools time to modify their curriculum and instructional practices to meet the needs of students before the first operational assessment. Field test items for Language Arts Literacy, Mathematics, and Science continued to be included with the GEPA 2000 – 2004 test administrations.

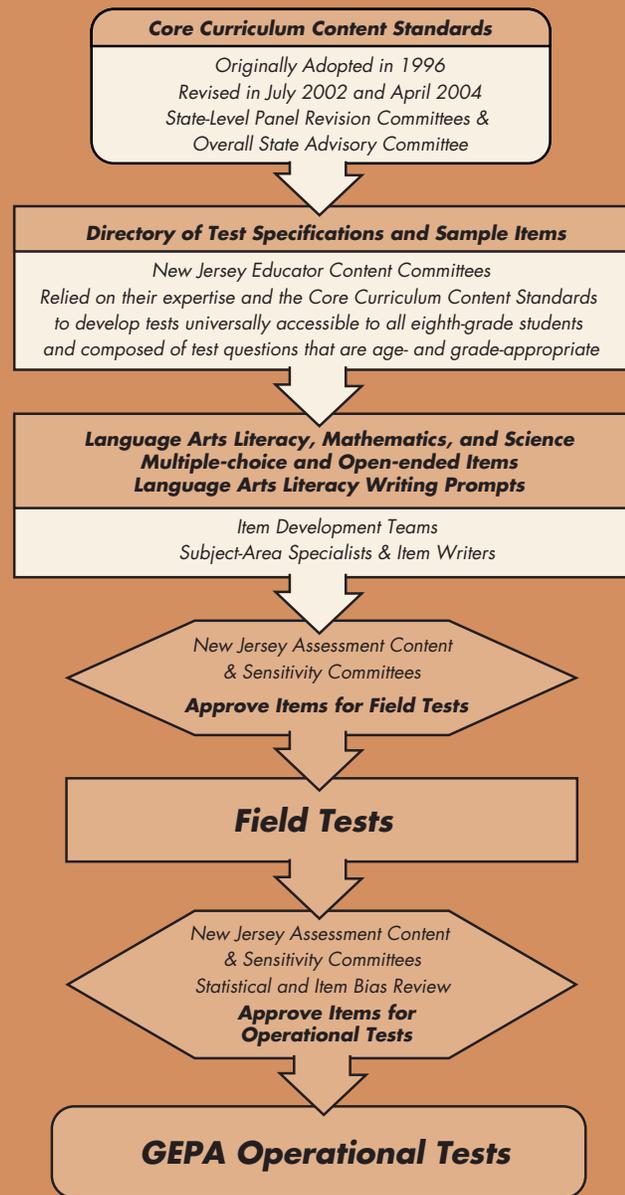
The operational test items included on the 2004 GEPA were aligned with the Core Curriculum Content Standards adopted in 1996. Since the operational test items on the 2005 GEPA will be aligned to the Core Curriculum Content Standards adopted by the Board in July 2002 and April 2004, the alignment of the 2004 field test items with both the 1996 Standards and the 2002/2004 Standards will be described in Section 2.4 of this test development chapter.

Following adoption of the original Core Curriculum Content Standards in 1996, the New Jersey Assessment Content Committees met through 1997 to develop a directory of test specifications and sample items for each content area to provide content/skill outlines and sample items. These directories describe the test, item formats, and test item scoring. This test specification work done by New Jersey educators serves as the foundation for all test item development.

The committees of New Jersey educators rely upon their expertise and the Core Curriculum Content Standards to design a test that is universally accessible to all eighth-grade students and is composed of test questions that are age- and grade-appropriate. The material in the three directories of test specifications and sample items is designed for use by curriculum specialists and teachers to improve instruction at the district, school, and classroom levels. Figure 2.1 summarizes the steps of the test development process beginning with the development of the Core Curriculum Content Standards and ending with an operational GEPA test form. Brief descriptions of the test content measured in Language Arts Literacy, Mathematics, and Science are presented in the following sections.

Figure 2.1

**GEPA Test Development Process**



*The Technical Advisory Committee (TAC) advises and assists the Office of Evaluation and Assessment in the development and implementation of the statewide testing program. TAC reviews and provides suggestions for each of the stages listed in the GEPA Test Development Process.*

### Language Arts Literacy

Language Arts Literacy measures students' achievements in reading and writing. Language Arts Literacy currently assesses knowledge and skills in two content clusters:

- Reading
- Writing

The Reading cluster consists of a narrative reading passage with ten multiple-choice and two open-ended items, and a persuasive reading passage with ten multiple-choice and two open-ended items. The passages are selected from published sources such as books, newspapers, magazines, and the Internet.

The Writing cluster for GEPA consists of three writing activities: a writing/persuade task in response to a prompt, a writing/speculate task in response to a picture, and a revise/edit task in response to a stimulus.

For an in-depth description of the Language Arts Literacy assessment, refer to the *Directory of Test Specifications and Sample Items for the Elementary School Proficiency Assessment (ESPA), Grade Eight Proficiency Assessment (GEPA), and High School Proficiency Assessment (HSPA) in Language Arts Literacy* (February 1998). The directory is available on-line at <http://www.njpep.org/assessment/TestSpecs/LangArts/TOC.html>, or by calling the New Jersey Department of Education, Publications Office, (609) 984-0549.

### Mathematics

Mathematics measures students' ability to solve problems by applying mathematical concepts. The GEPA Mathematics assessment measures knowledge and skills in four content clusters:

- Number Sense, Concepts, and Applications
- Spatial Sense and Geometry
- Data Analysis, Probability, Statistics, and Discrete Mathematics
- Patterns, Functions, and Algebra

Mathematics items are also classified and reported as Knowledge (requiring conceptual understanding or procedural knowledge) and Problem Solving (applying mathematical concepts). For the operational test, there are a total of 30 multiple-choice and 6 open-ended items in Mathematics. For an in-depth description of the GEPA Mathematics assessment, refer to the *Directory of Test Specifications and Sample Items for the Grade Eight Proficiency Assessment (GEPA) and the High School Proficiency Assessment (HSPA) in Mathematics* (February 1998). The directory is available on-line at <http://www.njpep.org/assessment/TestSpecs/MathTestSpec/GEPA Math/MathIndex.html>, or by calling the New Jersey Department of Education, Publications Office, (609) 984-0549.

## Science

Science measures knowledge and skills in three content clusters:

- Life Science
- Physical Science
- Earth Science

Science items are also classified and reported as Cognitive Skills (core knowledge, systems, history of science, and uses of technology) and Process Skills (problem-solving, mathematics tools, and selecting tools). For the operational test, there are a total of 45 multiple-choice and 3 open-ended items in Science.

For an in-depth description of the Science assessment, refer to the Directory of Test Specifications and Sample Items for the Grade Eight Proficiency Assessment (GEPA) and the High School Proficiency Assessment (HSPA) in Science, February 1998. The directory is available on-line at [http://www.njpep.org/assessment/TestSpecs/science\\_test\\_specs/Science\\_GEPA\\_HSPA/](http://www.njpep.org/assessment/TestSpecs/science_test_specs/Science_GEPA_HSPA/), or by calling the New Jersey Department of Education, Publications Office, (609) 984-0549.

Tables 2.1, 2.2, and 2.3 summarize the total points possible for Language Arts Literacy, Mathematics, and Science of the content areas of the operational GEPA administered in 2004.

### 2.2 Development of Test Items

The 2004 GEPA consists of two types of items:

- operational test items used to determine students' scores and
- field test items evaluated for use as future operational test items.

The 2004 operational test for Language Arts Literacy, Mathematics, and Science was composed of items field tested through 2003. The item development teams consisted of subject-area specialists and consulting item writers. These writers were teachers or former teachers with a great deal of specialized knowledge (e.g., education and training, years of classroom experience, familiarity with the student population, knowledge of the content area, and understanding of the pedagogy that defines the discipline) concerning their area of content expertise.

Each of the content areas consists of multiple-choice and open-ended items. The multiple-choice items are designed to measure those skills determined to be best measured by such item types, and the open-ended items are developed to measure those skills requiring students to do more than select a correct answer. That is, the open-ended items are designed to tap more complex and integrated skills. Language Arts Literacy includes a writing/persuade task, a writing/speculate task in response to a picture, and a revise/edit task.

The Measurement Incorporated/Pearson Educational Measurement item development process for each testing cycle begins with a formal review of the Core Curriculum Content Standards and the three directories of test specifications. Item-writing training sessions typically last from 8 to 16 hours over two days. The respective test development specialist for each content area conducts the training session. Between the first and second sessions, preliminary versions of test items developed in the first session are evaluated. At the second session, the training is focused on the items developed in the first session.

TABLE 2.1

**Total Points Possible for the Language Arts Literacy Component of the GEPA**

<b>Language Arts Literacy</b>			
<b>Total</b>	<b>62 points</b>		
Reading	36 points*		
Writing	26 points*		
Writing/Speculate	6 points*	1 – 6 points, ratings averaged	
Writing/Persuade	12 points*	1 – 6 points, ratings summed	
Revise/Edit	8 points*	0 – 4 points, ratings summed	
-----			
Interpreting Text	15 points*		
Analyzing/Critiquing Text	21 points*		

\*Cluster-level results show how students perform on the sets of items that measure particular knowledge and skills (clusters above the dotted line) or particular processes (clusters below the dotted line). Items in the Reading cluster contribute to the Reading knowledge and skills cluster as well as the Interpreting Text and Analyzing/Critiquing Text process clusters. However, each item counts only once in the total score.

Table 2.2

**Total Points Possible for the Mathematics Component of the GEPA**

<b>Mathematics</b>	
<b>Total</b>	<b>48 points</b>
Number Sense, Concepts, and Applications	12 points*
Spatial Sense and Geometry	12 points*
Data Analysis, Probability, Statistics, and Discrete Mathematics	12 points*
Patterns, Functions, and Algebra	12 points*
-----	
Knowledge	48 points*
Problem Solving	38 points*

\* Cluster-level results show how students perform on the sets of items that measure particular knowledge and skills (clusters above the dotted line) or particular processes (clusters below the dotted line). All Mathematics items are classified as Knowledge because all items require conceptual understanding or procedural knowledge. Some items also measure Problem Solving. Each Mathematics item counts only once in the total score.

Table 2.3

**Total Points Possible for the Science Component of the GEPA**

<b>Science</b>	
<b>Total</b>	<b>54 points</b>
Life	19 points*
Physical	19 points*
Earth	16 points*
-----	
Cognitive Skills	28 points*
Process Skills	26 points*

\* Cluster-level results show how students perform on the sets of items that measure particular knowledge and skills (clusters above the dotted line) or particular processes (clusters below the dotted line). Though an item on the GEPA can contribute to a cluster above the line (for example, Life) as well as a cluster below the dotted line (for example, Cognitive Skills), each item is counted only once in the total score.

At the training, each consulting item writer is asked to sign a Letter of Agreement. This letter specifies the confidentiality and security regulations. The agreement also outlines the ownership regulations. No confidential materials related to the project are released without explicit approval of the NJDOE Office of Evaluation and Assessment.

During the training, each item writer is given the following information:

- An overview of the GEPA
- Final test blueprint for each subject-area test and item specifications
- A description of the item formats used, including important characteristics of each format
- A description of the item writing process and measures to take to avoid writing biased items
- A listing of the security procedures followed during the item development process.

Important guidelines for the GEPA item development and test structure are outlined below.

1. Items are written to reflect what students know and understand based on classroom instruction and their mastery of skills included in the Core Curriculum Content Standards. Items are also designed to assess higher-order or critical thinking skills in varied contexts that students are likely to understand; yet, they are based upon solid theoretical frameworks.
2. For each content area, the multiple-choice items represent a range of difficulty. For example, approximately 25 percent of the items are relatively easy, 50 percent of the items are somewhat difficult, and 25 percent of the items are difficult. This range of difficulty provides for a distribution of items with p-values from approximately 0.30 to 0.95. This distribution allows for a range of difficulty that supports the established proficiency levels, yet is not so difficult that low-achieving students cannot be assessed adequately.
3. Item content for all of the items, including the writing-task prompts, is carefully reviewed to ensure that the items are free from gender, racial, ethnic and regional bias. Across all content areas of the GEPA and in any test material presented, there is a balance of gender and active/passive roles by gender.
4. Measurement Incorporated/Pearson Educational Measurement construct initial rubrics for each open-ended item in Language Arts Literacy, Mathematics, and Science.
5. Writing task prompts for Language Arts Literacy are written in such a way that they focus on experiences that eighth-grade students may have every day. However, care must be taken to ensure that the writing task prompts are not intrusive in nature and do not elicit personal information of a biographical, religious, political, or affective nature. Topics must be chosen so that no group of eighth-grade students is put at a subject-related disadvantage. Instead, each writing task prompt is designed to sample the skills and abilities demanded of eighth-grade students. Each writing task is developmentally appropriate for students in both the academic and nonacademic environments.

As items are developed, Measurement Incorporated/ Pearson Educational Measurement document each item's relevancy to the Core Curriculum Content Standards and to the directories of test specifications. During this process, each item is assigned a unique item ID number or coding system number. This unique number identifies the following: content area, skill measured, standard, and associated materials such as a reading passage, artwork, or display of data. The number is used to track the item throughout the development process and its eventual use on the operational test.

All items prepared by item writers are reviewed, revised, and edited by the subject area specialists and editors prior to review by the New Jersey Assessment Content Review Committees. Also, the New Jersey Assessment Sensitivity Review Committee approves passages used on the Language Arts Literacy section.

In preparation for the 2004 field items, a total of 192 Language Arts Literacy, 144 Mathematics, and 132 Science items were requested by the NJDOE Office of Evaluation and Assessment staff. Table 2.4 shows the number of multiple-choice and open-ended items specified for each content area.

**TABLE 2.4**  
**Development Goals for the 2004 Field Test Items**

	<i>Multiple-choice Items</i>	<i>Open-ended Items</i>	<i>Total Items</i>
	<b>Goal</b>	<b>Goal</b>	<b>Goal</b>
<b>Language Arts Literacy</b>	160	32	192
<b>Mathematics</b>	120	24	144
<b>Science</b>	120	12	132
<b>TOTAL</b>	<b>400</b>	<b>68</b>	<b>468</b>

### 2.3 Item Review Process

The New Jersey Assessment Content Committee members provide expert judgments on the alignment of each test item with the Core Curriculum Content Standards and the content-specific test specifications. The committee members represent school districts across all District Factor Groups. Table 2.5 shows the District Factor Groups represented on each of the Content and Sensitivity Committees.

Prior to field testing, all items are reviewed by the Office of Evaluation and Assessment staff, the Content Committees, and the Sensitivity Committee during item review meetings. Each test item is reviewed to determine if the item meets test specifications and addresses an appropriate level of difficulty. Committees also ensure that test questions are not offensive and do not reinforce negative stereotypes, and that test questions appropriately reflect multicultural society.

Figure 2.2 presents a sample of the form that must be marked “Definitely Use” or “Revise and Use With Approval” during review committee meetings before an item is included in a field test. The percentage of test items accepted for field testing depends on the content area and the item type. The range of acceptance generally is 60-80% at this item review stage. During review, committee members approve items, amend or revise items, or reject items.

**TABLE 2.5**

**District Factor Groups (DFG) Represented on the GEPA Content and Sensitivity Committees**

<b>DFG</b>	<b>Language Arts Literacy</b>	<b>Mathematics</b>	<b>Science</b>	<b>Sensitivity</b>	<b>Total</b>
A	1	1	3	1	6
B	3	2	0	1	6
CD	1	1	0	0	2
DE	4	2	1	0	7
FG	3	3	0	2	8
GH	2	4	3	0	9
I	1	0	4	0	5
J	0	1	1	0	2
Retirees	3	6	3	9	21
Private School	0	0	1	0	1
Not in Districts	0	2	1	1	4
<b>Total</b>	<b>18</b>	<b>22</b>	<b>17</b>	<b>14</b>	<b>71</b>

Committee members sign a Confidentiality and Security Agreement noting they must maintain the security of the testing materials by not discussing and disclosing any confidential information related to the program.

**FIGURE 2.2**

**Item Approval Before Field Test**

<b>Sensitivity</b>			<b>Content</b>		
<b>*Comments</b>			<b>*Comments</b>		
Sensitivity Issue	Yes	No	Meets Specifications	Yes	No
<i>If yes, identify category and explain*</i>			Appropriate Difficulty	Yes	No
			Accurate Coding	Yes	No
Definitely Use			Definitely Use		
Revise and Use With Approval			Revise and Use With Approval		
Revise and Resubmit			Revise and Resubmit		
Do Not Use*			Do Not Use*		

Sensitivity Sign-off

Date

Content Chairperson's Signature

Date

TABLE 2.6

**GEPA 2004 Content and Sensitivity Committee Meetings**

<b>Language Arts Literacy Committee</b>
LAL Passage Review Tues – Thu, April 27 – 29
1st Item Review Tue – Fri, June 22 – 25
2nd Item Review Mon – Fri, August 2 – 6
3rd Item Review Tue – Wed, September 14 – 15
Statistical Item Review Mon – Fri, August 9 – 13
<b>Mathematics Committee</b>
1st Item Review Tue – Thu, June 22 – 24
2nd Item Review Wed – Fri, August 4 – 6
Statistical Item Review Mon – Thu, August 9 – 12
<b>Science Committee</b>
1st Item Review Tue – Wed, June 22 – 23
2nd Item Review Mon – Tue, August 2 – 3
Statistical Item Review Mon – Tue, August 9 – 10
<b>Sensitivity Committee</b>
LAL Passage Review Wed – Thu, April 28 – 29
Statistical Item Review Mon, August 9

The committees also meet to review item statistics of the field test items. Committee meetings during the spring and summer of 2004 are listed in Table 2.6. Committee members reviewed field test item statistics during meetings in early August 2004.

At the statistical review, committee members consider how well students did on the each field test question in comparison to the other questions on the GEPA. If an item yields good statistics, it will become part of the operational pool for future GEPA tests. Otherwise, it will be eliminated or revised and re-field tested.

Prior to field testing, the field tested open-ended items and writing prompts must go through rangefinding to determine the scores on sample student responses. The field test rangefinding process involves scoring 30 student responses for each of the open-ended items and writing prompts. These 30 responses are selected to represent the wide range of responses to that item. The papers are scored by one or two content committee members, the NJDOE Content Coordinator, and representatives from Measurement Incorporated.

In Language Arts Literacy, the responses are scored according to the generic rubric for either reading or writing as appropriate. Use of these generic rubrics ensures that student responses are scored in the same way for the demonstration of the same level of knowledge and skills regardless of the prompt or the year.

For Mathematics and Science, each item has a unique scoring rubric, based on the generic one for each area. During rangefinding, the item specific rubric is refined, if necessary, to define each score point clearly. The rangefinding process aids in delineating between a 0 & 1, 1 & 2, and a 2 & 3 score point response. The holistic scoring guide is used quite often to refine the tenuous line between the score points.

For all content areas, the scored field test responses and the rubrics are used to create the holistic scoring guide, which is used to help refine the lines between the score points. This guide is then used to train the scorers of that item. If there is any problem or question with the scoring of a student's response, the NJDOE Content Coordinator is contacted and makes a final decision for the score of that paper. After the open-ended papers have been scored, the scorers discuss the types of responses and problems, if any, found during scoring of each item. The scor-

ing director then writes a brief summary of these comments and sends it, along with a copy of each item, rubric, sample answer and rangefinding paper to the statistics review. Other than this packet, the same field test review procedures are used for the open-ended items and the multiple-choice items.

Pearson Educational Measurement computes item means, response frequencies, biserial correlations (with operational test total scores), and other descriptive statistics. Prior to the presentation of items and statistics to reviewers, the NJDOE Office of Evaluation and Assessment defined boundaries within which item statistics should fall. In general, items with p-values below 0.30 or above 0.95 were considered usable only if a strong content argument could be made for their inclusion in the item bank. An item could be flagged for low or high p-value and/or low biserial correlation with operational test total scores.

For the statistical item review, the Mantel-Haenszel statistic is calculated to show whether or not students are responding to an item in a way that their overall ability (as measured by the operational test) would lead us to expect. This statistic takes into consideration both group membership (by race or by gender) and ability. The Mantel-Haenszel statistic is used for a classification determination of category A, B, or C. An item in Category A shows no or minor relationship between group membership and performance. Category B items are somewhat suspect. Category C items show a substantial relationship between group membership and item performance and must be examined carefully by the committees to make sure these items are not biased. The Mantel-Haenszel statistic is used at Educational Testing Service (ETS) as a classification determination of category A, B, and C as described by Zieky (1993):

---

*Category A) MH D-DIF not significantly different from zero*

*OR*

*absolute value less than 1.0*

*Category B) MH D-DIF significantly different from zero and absolute value of at least 1.0*

*AND EITHER*

*(1) less than 1.5*

*OR*

*(2) not significantly greater than 1.0*

*Category C) MH D-DIF significantly greater than 1.0*

*AND*

*absolute value 1.5 or more. (p. 342)*

---

For every open-ended item and writing prompt, the Sensitivity Committee reviews frequency distributions for the range of scores of the following student groups: total, white, African-American, Hispanic, Asian, American-Indian, male, and female.

**FIGURE 2.3**

**Item Approval Before Operational Test**

<b>Sensitivity</b>		<b>Content</b>	
*Comments		*Comments	
Sensitivity Issue <input type="checkbox"/> Yes <input type="checkbox"/> No		Appropriate Difficulty <input type="checkbox"/> Yes <input type="checkbox"/> No	
If yes, identify category and explain*		PVal = Biserial =	
Mantel-Haenszel Category C <input type="checkbox"/> W-AA <input type="checkbox"/> W-H <input type="checkbox"/> M-F			
<input type="checkbox"/> Yes <input type="checkbox"/> No	Definitely Use		<input type="checkbox"/> Yes <input type="checkbox"/> No
<input type="checkbox"/> Yes <input type="checkbox"/> No	Revise and Use With Approval**		<input type="checkbox"/> Yes <input type="checkbox"/> No
<input type="checkbox"/> Yes <input type="checkbox"/> No	Revise and Re-Field Test		<input type="checkbox"/> Yes <input type="checkbox"/> No
<input type="checkbox"/> Yes <input type="checkbox"/> No	Do Not Use*		<input type="checkbox"/> Yes <input type="checkbox"/> No

Sensitivity Sign-off

Date

Content Chairperson's Signature

Date

\*\*Requires director's approval

For the multiple-choice items field tested during 2004, nine items in Language Arts Literacy, eight items in Mathematics, and four items in Science were flagged. For each of the 21 flagged multiple-choice items, the Sensitivity Committee marked the “No” box indicating that they did not determine a sensitivity issue. However, the content committees rejected one of the multiple-choice flagged items in Language Arts Literacy, one of the multiple-choice flagged items in Mathematics, and one of the multiple-choice flagged items in Science.

Figure 2.3 presents a sample of the form that must be marked “Definitely Use” or “Revise and Use With Approval” during review committee meetings of the field test statistics before an item is included on an operational base test.

**TABLE 2.7**  
**LANGUAGE ARTS LITERACY - READING**  
**Number of Field Test Items Approved During Statistical Review**

Passages	Field-Tested		Approved		Revise & Re-Field Test		Do Not Use	
	MC	OE	MC	OE	MC	OE	MC	OE
Narrative 1	20	4	0	0	0	0	20	4
2*	20	4	16	0	0	0	4	4
3	20	4	17	2	0	0	3	2
4	20	4	16	3	0	0	4	1
Persuasive 1	20	4	0	0	0	0	20	4
2	20	4	18	3	0	0	2	1
3*	20	4	19	1	0	0	1	3
4	20	4	14	0	0	0	6	4
<b>TOTAL</b>	<b>160</b>	<b>32</b>	<b>100</b>	<b>9</b>	<b>0</b>	<b>0</b>	<b>60</b>	<b>23</b>

\* Following the statistical review, committee members wrote new open-ended items for Narrative Passage 2 and Persuasive Passage 3. These passages will be refield-tested with these new open-ended items.

Tables 2.7 – 2.11 present the number of items field tested during the administration.

Table 2.7 shows 160 multiple-choice items and 32 open-ended items were field tested for the Reading component of Language Arts Literacy, which includes four narrative passages and four persuasive passages. During the statistical review, the Language Arts Literacy committee determined that only Narrative Passage 3, Narrative Passage 4, and Persuasive Passage 2 could be approved for the operational test. Because the committee agreed that many of the multiple-choice items for Narrative Passage 2 and Persuasive Passage 3 performed very well, committee members revised and approved open-ended items for these two passages. Narrative Passage 2 and Persuasive Passage 3 will be refield tested with the new open-ended items.

For the Writing component of the Language Arts Literacy, two persuasive prompts and two revise/edit tasks were field tested. The two persuasive prompts and two revise/edit tasks were approved for the operational test during the statistical review. However, with the adoption of the new test specifications, the revise/edit tasks will no longer be included in the writing component beginning with the GEPA 2005 test administration.

Table 2.8 reports the results by content cluster for the 140 multiple-choice items and 23 open-ended Mathematics items were field tested in 2004. Each content cluster is further divided into macros. The macros are listed in the *Directory of Test Specifications and Sample Items for the Grade Eight Proficiency Assessment (GEPA) and the High School Proficiency Assessment (HSPA) in Mathematics* (February 1998). Table 2.8 indicates that 70.7% Mathematics multiple-choice items and 43.5% Mathematics open-ended items were approved for an operational base test.

Because the GEPA operational test beginning with the 2005 administration will align with the test specifications for the 2002/2004 Core Curriculum Content Standards, the Mathematics items approved for the operational test use also are classified according to these new test specifications. Results of this classification are summarized in Table 2.9. Additional information about the new test specifications including the associated strands is located at <http://www.njpep.org/assessment/TestSpecs/MathTestSpec/GEPAMath/Macros.html>.

Table 2.10 shows that 180 multiple-choice and 11 open-ended Science items were field tested in 2004. This indicates that 91.1% Science multiple-choice items and 100% Science open-ended items were approved for an operational test. The number of Science items field tested for each content cluster as well as by cognitive skill and process skill is shown in the table.

Because the GEPA 2005 administration will align with the science test specifications for the 2002/2004 Core Curriculum Content Standards, the Science items approved for the operational test use are classified according to the new test specifications. The Science committee realigned the field test items to the new test specifications. Results of this classification are summarized in Table 2.11. Two items originally accepted according to the earlier test specifications did not “map” to the new specifications. Therefore, these test items will not appear on an operational test. Additional information about the new science test specifications is located at [http://www.njpep.org/assessment/TestSpecs/ScienceGEPA/TestSpecsRev9\\_04.doc](http://www.njpep.org/assessment/TestSpecs/ScienceGEPA/TestSpecsRev9_04.doc).

TABLE 2.8

## MATHEMATICS

## Number of Field Test Items Approved During Statistical Review

Content Cluster	Macro	Field-Tested		Approved		Revise & Re-Field Test		Do Not Use	
		MC	OE	MC	OE	MC	OE	MC	OE
Number Sense, Concepts and Applications	A	13	1	11	0	0	0	2	1
	B	14	0	12	0	1	0	1	0
	C	9	3	6	1	1	2	2	0
Spatial Sense and Geometry	A	12	1	9	0	0	0	3	1
	B	6	3	5	2	0	0	1	1
	C	14	3	6	1	3	0	5	2
Data Analysis, Probability, Statistics, and Discrete Mathematics	A	14	3	8	2	1	0	5	1
	B	13	1	8	0	4	0	1	1
	C	9	2	6	1	0	0	3	1
	D	6	1	4	0	0	0	2	1
Patterns, Functions, and Algebra	A	16	3	13	2	0	1	3	0
	B	14	2	11	1	1	0	2	1
<b>TOTAL</b>		<b>140</b>	<b>23</b>	<b>99</b>	<b>10</b>	<b>11</b>	<b>3</b>	<b>30</b>	<b>10</b>

TABLE 2.9

## MATHEMATICS

## Number of Field Test Items Approved When Classified with 2002/2004 Core Curriculum Content Standards

Cluster	Strand	Number of Items	
		MC	OE
Number and Numerical Operations	A	9	–
	B	10	1
	C	11	–
Geometry and Measurement	A	13	–
	B	4	–
	C	–	1
	D	1	–
	E	3	2
Patterns and Algebra	A	9	–
	B	–	–
	C	10	3
	D	6	–
Data Analysis, Probability, and Discrete Mathematics	A	8	–
	B	8	2
	C	6	1
	D	1	–
<b>TOTAL</b>		<b>99</b>	<b>10</b>

TABLE 2.10

## SCIENCE

## Number of Field Test Items Approved During Statistical Review

Content Skill	Field-Tested		Approved		Revise & Re-Field Test		Do Not Use	
	MC	OE	MC	OE	MC	OE	MC	OE
<b>Life</b>								
Cognitive	23	3	20	3	2	0	1	0
Process	31	1	28	1	0	0	3	0
<b>Physical</b>								
Cognitive	33	0	31	0	0	0	2	0
Process	54	3	49	3	1	0	4	0
<b>Earth</b>								
Cognitive	12	1	11	1	1	0	0	0
Process	27	3	25	3	0	0	2	0
<b>TOTAL</b>	<b>180</b>	<b>11</b>	<b>164</b>	<b>11</b>	<b>4</b>	<b>0</b>	<b>12</b>	<b>0</b>

TABLE 2.11

## SCIENCE

## Number of Field Test Items Approved When Classified with 2002/2004 Core Curriculum Content Standards

Cluster	Skill	Knowledge		Application	
		Number of Items		Number of Items	
		MC	OE	MC	OE
Life		11	—	37	4
Physical		15	—	64	3
Earth		9	—	26	4
<b>TOTAL</b>		<b>35</b>	<b>---</b>	<b>127</b>	<b>11</b>

## 2.4 Operational Test Development

Following the 1998 through 2001 administrations, GEPA examiners completed a feedback form seeking suggestions and concerns related to the testing procedures. Questions related to timing, directions, and answer documents were asked specifically for each content area tested. Also, examiners were asked to identify questions that arose on issues and topics not addressed in the test booklets, directions, or coordinator or examiner manuals.

A sample of the 2001 questions is provided below:

- Was the time allotted for students to complete the test sufficient?
  - too much time
  - time about right
  - too little time
- Were the directions clear?
  - yes, directions were clear
  - no, directions were somewhat confusing
- Was the space provided for student responses in the answer folder sufficient?
  - adequate space
  - not enough space

Information from the examiners' responses assisted the Office of Evaluation and Assessment with determining the operational testing procedures.

The GEPA Content Committees assisted with recommending the emphases and priorities reflected in the number of items for each item type and cluster on the operational test. The operational test specifications appear in Table 2.12.

**TABLE 2.12**  
**Operational Test Specifications**

Content Areas	Cluster	Number of Items		
		MC	OE	Total
Language Arts Literacy	Reading	20	7	27
	Writing	20	4	24
	Writing/Speculate		1	1
	Writing/Persuade		1	1
	Revise/Edit		1	1
Mathematics	Number Sense, Concepts, and Applications	30	6	36
	Spatial Sense and Geometry	6	2	8
	Data Analysis, Probability, Statistics, and Discrete Mathematics	9	1	10
	Patterns, Functions, and Algebra	6	2	8
Science	Life	9	1	10
	Physical	16	1	17
	Earth	16	1	17
	Earth	13	1	14

Following the approval of test items for use on operational tests by the Content and Sensitivity Review Committees, Assessment and Evaluation Services (AES) selected items for each GEPA administration to meet test specifications for Language Arts Literacy, Mathematics, and Science.

Relevant considerations for operational test development included content quality and scope, cluster representation, and appropriate item difficulty indices. The new operational test was parallel to the content, format, and statistical characteristics of the previous operational forms. Selecting test items for the operational tests is an iterative process to create test forms that are the perfect combination of content and statistical information. Through the iterative process, item content took precedence over statistical characteristics.

The operational test development used the Rasch model to pre-equate cluster and total test scores. Rasch item difficulty statistics were calibrated to the previous test administration. Common items were chosen to link the Mathematics and Science operational tests to previous forms for equating purposes. For Language Arts Literacy, the forward and backward items for equating purposes were specified. For each operational test, AES produces a spreadsheet that includes the following information for both the previous operational test and newly developed operational test.

- Item identifier with item type (multiple-choice or open-ended), content clusters, and skill clusters
- Common items for equating
- P-values and biserial correlations
- Item difficulties with sums and averages for clusters and total test

### **2.5 Review and Approve Operational Test Forms**

The Office of Evaluation and Assessment approved the operational test forms for each GEPA administration. AES and PEM assisted with quality control that included:

- Confirm that each test item appears on the operational test as it was approved by the Content and Sensitivity Review Committees.
- Confirm that all test specification requirements are met.
- Check adequacy of common item set (i.e., in terms of size, content and skill representation)
- Double-check that the item and mean difficulty levels are accurate and meet requirements.
- Take the test to be certain all content considerations including content/skill/topic balance, correct keys, no clueing, and correct graphics are met.

### **2.6 Test Materials for Visually Impaired Students**

The Office of Evaluation and Assessment works with the New Jersey Commission for The Blind and Visually Impaired to identify items with graphs, charts, and illustrations that may not translate well into Braille or large-print versions of the test. For 2004, the Writing/Speculate prompt from Language Arts Literacy, six items from Mathematics, and three items from Science were removed from the Braille form. Also, one Science item was removed from the large-print form.

## **CHAPTER 3: TEST ADMINISTRATION**

### **3.1 Participation**

In 1988, the New Jersey State Legislature passed a law (18A:7C-6.2) requiring that a test be given to all eighth-grade students in public schools in New Jersey to assess their progress toward mastering the skills they will need to graduate from high school. All eighth-grade public school students must take the GEPA. This includes:

- General education students
- Limited-English Proficient (LEP) students
- Special Education (SE) students
- Students with Disabilities (Section 504)

In accordance with the Individuals with Disabilities Education Act (IDEA), students who are receiving special education services must participate in each subject area of the age-appropriate statewide assessment with the following exception:

Students with disabilities shall participate in the Alternate Proficiency Assessment in each content area where the nature of the student's disability is so severe that the student is not receiving instruction in any of the knowledge and skills measured by the general statewide assessment and the student cannot complete any of the types of questions on the assessment in the content area(s) even with accommodation and modifications. (New Jersey Administrative Code Chapter 6A:14-4.11[a]2)

The Alternate Proficiency Assessment (APA) is a portfolio-style assessment designed to measure progress toward achieving New Jersey's state educational standards for those students with severe disabilities who are unable to participate in the New Jersey Assessment of Skills and Knowledge (NJASK), the Grade Eight Proficiency Assessment (GEPA), or the High School Proficiency Assessment (HSPA).

### 3.2 Test Security Procedures

The test booklet and its contents are secure materials. They are not to be read or copied, either wholly or in part, for any purpose without express written permission from the New Jersey Department of Education. It is the responsibility of the school districts to guarantee the security of the test materials. Security breaches may have financial consequences for the district, professional consequences for staff, and disciplinary consequences for students.

The items and passages contained in the test booklet must remain confidential because some test items will reappear in future versions of the tests. The answer folders (approximately 56 pages) contain grids for marking the answers to multiple-choice questions. Also, the answer folders are used by students for writing responses to the open-ended questions, and the writing essay prompts. Some items and passages included in the Group 1 answer folders are secure and also must be kept confidential for future testing. The security of test items and passages is required to maintain the stability of the test item pool over time from a technical perspective and to enable comparisons to be made from one year to the next. Examiners, proctors, and other school personnel are prohibited from discussing or disclosing any test items before, during, or after the test administration.

The following are secure materials for the administration:

- Test booklets
- Used answer folders and all used/unused Group 1 answer folders
- All other answer folders until after testing
- Mathematics Reference Sheets until after testing

Pearson Educational Measurement (PEM) assigns a unique identification number to each secure test booklet and answer folder. The unique identification numbers are listed on security checklists. The unique identification number appears as a bar-code on test booklets and Group 1 answer folders. Following the test administration, PEM compares bar-code scan files of returned test booklets and answer folders with distribution files to determine if all secure materials have been returned from each school and district. PEM contacts any district with missing secure test booklets or answer folders. For the 2004 administration, PEM scanned more than 150,000 secure test booklets and answer folders.

The NJDOE Office of Evaluation and Assessment outlined security procedures in the *Test Manual*. District test coordinators were trained in these procedures during regional meetings held by the Office of Evaluation and Assessment in January 2004. The district test coordinators' training and the *Test Manual* included responsibility descriptions for the district test coordinator, school test coordinator, and examiner.

1. The chief school administrator or designee must sign for the initial shipment of test materials after presenting the Authorization to Receive Secure Test Materials form to the agent of the delivery service when the materials are delivered.
2. When not being used during testing, test materials must be stored in a secure, locked place that is accessible only to individuals whose access has been authorized by the school test coordinator. During testing, secure materials must not be removed from the testing room for review or photocopying. **Security of test materials must be maintained at all times.**
3. Each test booklet and answer folder has a **unique identification number**. Students must use the same test booklet and the same answer folder for each day of testing. On the first day of testing, students should print their name on the front cover of the test booklet assigned to them, and record the number and form letter of that test booklet on their answer folder.
4. Teachers are NOT to be given their own test booklet. The shrink-wrapped packaging on the test booklets may be opened for distribution just prior to testing.
5. Each day's section of the test booklet is sealed on all open sides. There are separate seals for the Science section, the Mathematics section, and Day 1 and Day 2 of the Language Arts Literacy section of the test. These seals must not be broken until the student breaks them the day that test section is administered.
6. District and school test coordinators must use the District and School Security Checklists to maintain an accurate record of the chain of distribution and collection of all test booklets.
7. Answer folders must not be duplicated or handscored.
8. An answer folder must be gridded for every enrolled Grade 8 student regardless of APA/Exemption status.
9. An Irregularity Report form is used to report irregularities involving test booklets, answer folders, or anything that could impact test takers.
10. The principal and the chief school administrator or his/her designee must review and sign the completed Header sheets before they are submitted for scoring. The signatures affirm that the number of answer folders returned is correct and that all GEPA test administration procedures outlined in the manuals have been followed.
11. The Office of Evaluation and Assessment, in cooperation with county offices, monitors all aspects of testing and the implementation of security procedures at selected sites. Announcements of security visits are not made in advance.

Breach test forms and examiner's manuals were prepared in the event of a security breach. In schools with the security breaches, appropriate staff members completed each student's name, date of birth, and answer folder number so that the alternate scoring can occur properly for the students. Specialized scoring and reporting included developing alternate test score keys, conversion tables, and reports.

### 3.3 Test Administration Procedures

The district test coordinators, school test coordinators, and examiners are responsible for the proper administration of the test. The district test coordinator is responsible for ensuring that examiners are selected and trained. All examiners must be certified teachers currently employed by the school. The district and school test coordinators, and examiners must read the *Test Manual* and *Examiner's Manual* carefully to get an overview of all activities.

Student Rosters with appropriate Special Codes must be prepared to include each and every eighth-grade student in the district. The information from the rosters is used to code the “School Use Only” section of the student information grid on page one of the answer folder; to verify the pre-ID label, if applicable.

The Student Rosters must:

- List each eighth-grade student's name, date of birth, gender, and ethnicity
- Identify students with SE classifications, IEP exemptions/accommodations, or Section 504 status
- Identify students who are designated Title 1, economically disadvantaged, Limited English Proficient, and/or migrant status
- Designate coding for student's time in district/time in school less than one year

Information from the Student Rosters is used to:

- ensure students are testing in the correct room
- code the “School Use Only” section of the student information grid on the answer folder
- verify correct gridding by students, and to
- verify that correct data appears on the pre-ID label for districts using labels.

Test booklets and answer folders are distributed to examiners only on the morning of each day of the test administration. Specific instructions for the test administration are contained in the *Examiner's Manual*. The examiners' familiarity with the materials and the prescribed procedures is essential to the successful administration of the test. During the examiners' training, district and school test coordinators emphasize that students can be given no assistance or coaching beyond what is specified in the manual.

When more than 25 students are tested in one room, the examiner uses the assistance of proctors. The school test coordinator briefs the proctors on the test materials and procedures, and specifies their responsibilities before, during, and after test administration. Proctors help in distributing and collecting non-secure materials, in observing students from different points in the room during test administration, and in answering student questions when there is a problem related to the test directions.

Total testing time (including time for distributing and collecting materials, reading directions, and taking breaks) is approximately nine hours over four successive days. The GEPA test administration must be scheduled in the morning. The Science, Mathematics, and Language Arts Literacy content-area tests were administered on the specified dates during the regular and make-up testing weeks.

### 3.4 Test Accommodations

To ensure that students are tested under appropriate conditions, the Department of Education has adopted test accommodations and modifications that may be used when testing special populations of students. The content of the test typically remains the same, but administration procedures, setting, and answer modes may be adapted. Students requiring accommodations must be tested in a separate location from general education students.

**General education students** receive no special testing accommodations other than the standard room setup and materials distribution described in the examiner's section of the *Test Manual*.

**Limited English Proficient (LEP)** students are tested with one or more of these accommodations:

- Additional time up to 150% of the administration times indicated.
- Translation of directions only to the student's native language. Translations of passages, items, prompts, and tasks are NOT permitted.
- Use of a bilingual dictionary, preferably one normally used by the student as part of the instructional program

**Special education (SE) students** must take the GEPA unless their Individualized Education Program (IEP) specifically exempts them. These IEP exempt students must then take the Alternate Proficiency Assessment.

Students with disabilities eligible for special education and related services and those students eligible under Section 504 of the Rehabilitation Act may have accommodations and/or modifications during administration of the statewide assessment.

Any accommodations or modifications of test administration procedures for students eligible for special education under IDEA or eligible under Section 504 of the Rehabilitation Act of 1973 must be specified in the student's IEP or 504 accommodation plan. Accommodations or modifications must be consistent with the instruction and assessment procedures used in the student's classroom. Students eligible for modifications under Section 504 may not be classified but do have a permanent or temporary impairment in a major life function (for example: performing manual tasks, walking, seeing, hearing, speaking, etc.).

**Visually impaired students** may take either a Braille or large-print version of the test. Specific instructions for administering the Braille and large-print versions of the test are provided in the supplementary instructions for examiners administering these forms.

Students using the Braille test booklets:

- are instructed to bring a Braille ruler and a talking calculator to the test session.
- are instructed to skip some items identified in the Braille instructions. The spaces for these items must be left blank on the student answer folder.
- have answer folders transcribed from Braille version by the examiner.
- dictate their answers to the examiner or use a device that produces Braille. For dictations and responses recorded in Braille:
  - Students must indicate all punctuation and must spell all key words.
  - Examiners must transcribe the Brailled responses into the regular answer folder.

Students using the large-print test booklets:

- mark their answers in the large-print answer folders.
- may be instructed to skip some questions. The spaces for these questions must be left blank in the student's large-print answer folder.
- who dictate responses on open-ended items and writing tasks indicate all punctuation and spell key words.

Accommodations and modifications of test administration procedures for the statewide assessments are listed in the *Test Manual*, the *Examiner's Manual*, and at <http://www.state.nj.us/njded/specialed/accom900.htm>.

If a student requires an accommodation or modification that is not listed, district staff are instructed to contact the Office of Evaluation and Assessment, GEPA Coordinator. Accommodations or modifications must be recorded on the student's answer folder by codes (A, B, C, or D). Accommodations or modifications are classified as follows:

- A = Setting Accommodations
- B = Scheduling Accommodations
- C = Test Materials/Modifications
- D = Test Procedures Modifications

## CHAPTER 4: SCORING

### 4.1 Multiple-choice Items

Each multiple-choice item contributes one point to the total raw score for each content-area test. Responses for multiple-choice items are machine scored. The score points of multiple-choice items received for a content area are the total number of multiple-choice items answered correctly. For the Mathematics and Science content areas and the Language Arts reading component, the total score points of multiple-choice items are combined with the total number of points from the open-ended items for a student's score. For Language Arts Literacy, the reading component score points are added to score points received from the open-ended scoring of the three writing tasks which compose the writing component.

### 4.2 Open-ended Items

During April and May of 2004, Measurement Incorporated (MI) under subcontract to Pearson Educational Measurement (PEM) scored the student writing responses, and the reading, mathematics, and science open-ended items. MI has a staff of highly-trained scorers who must have at least a bachelor's degree and who must undergo rigorous and ongoing training and monitoring during the scoring process. Each open-ended item and each writing prompt was read independently by two scorers. If the two scorers disagreed by more than one point, a third scorer evaluated the response. Appendix A presents information about how the three scores are resolved for each of the content areas.

Table 4.1 shows the number of writing responses and open-ended items scored for the operational test.

**TABLE 4.1**

### **Number of Writing Prompt and Open-ended Items Scored**

<b>Content Area</b>	<b>Number of Writing Prompt and Open-ended Items Scored</b>
Language Arts Literacy	1,558,449
Reading	888,542
Writing	669,907
Speculate	222,331
Persuade	222,830
Revise/Edit	224,746
Mathematics	1,332,779
Science	667,275
<b>TOTAL</b>	<b>3,558,503</b>

### **Scorer Selection**

Many individuals are responsible for ensuring the successful scoring of any large-scale assessment such as the GEPA. Key to the process of scoring the responses accurately and reliably are MI's senior project managers, scoring directors, team leaders, the scorers, and clerical aides.

MI's senior project managers work closely with Content Coordinators in the Office of Evaluation and Assessment. Current procedures for scoring the GEPA open-ended and writing responses are consistent with those used since the inception of a performance-based writing component in the New Jersey statewide assessment. Scoring of the open-ended and writing responses is monitored by trained, experienced personnel who have met the same rigorous standards established with the initial holistic scoring study conducted in 1986.

For selecting team leaders, MI's management staff and scoring directors reviewed the files of all returning staff who have previously scored the GEPA. The MI staff looked for people who were experienced team leaders with a record of good performance on previous projects and also considered scorers who have been recommended for promotion to the team leader position.

Many of the MI scorers have repeatedly scored the GEPA for previous test administrations. MI's procedures for selecting new scorers are very thorough. After advertising in local newspapers, with the job service, and elsewhere, and receiving applications, staff in MI's human resources department review applications and schedule interviews for qualified applicants. Qualified applicants are those with a four-year college degree in English, language arts, education, mathematics, science, or a related field. Each qualified applicant must pass an interview by experienced MI staff, write an acceptable essay, and receive good recommendations from references. All the information about each applicant is reviewed before offering employment.

MI is an equal opportunity employer that actively recruits minority staff. Historically, their temporary staff on major projects averages about 70 percent female, 30 percent male, 76 percent Caucasian, and 24 percent minority. MI strongly opposes illegal discrimination against any employee or applicant for employment with respect to hire, tenure, terms, conditions, or privileges of employment, or any matter directly or indirectly related to employment because of race, color, religion, sex, age, handicap, national origin, or ancestry.

### **Rangefinding**

Rangefinding is one of the most important elements of the scoring process. Rangefinding meetings provide an opportunity for finalizing scoring rubrics (in content areas with specific item rubrics) and making scoring decisions and interpretations regarding scoring issues before team leader and scorers' training begins. (See Appendix A for rubrics.) It is important that as many of the item-specific problems as possible be resolved prior to scorers' training so that scoring decisions can be made during scoring.

After consulting with PEM to determine when the first “live” student responses would be available, MI scheduled a rangefinding meeting in Durham, other MI sites (operational test), and New Jersey (field test) to establish “true” scores for a representative sample of open-ended items. At this meeting, Office of Evaluation and Assessment staff members, content committee members, and the MI project leaders read and scored 60-225 responses, which exemplified various points of the rubric and score scale. The number of responses varied according to the content area and score scale. The responses were selected from a broad range of New Jersey school districts in order to ensure that the sample was representative of overall student performance. Rangefinding took from two to six days per content area, depending on the number of items tested.

### **Development of Scoring Guides**

After the rangefinding responses were discussed and received a final score, MI used the selected responses to develop scoring guides, training sets (practice papers) and/or qualifying sets for each content area. Scoring guides consisted of three or more examples of each score point in score point order. In some content areas, the papers were annotated. Training and qualifying sets were clearly anchored papers in random score point order. Sufficient copies were made so that all scoring directors, team leaders, and scorers had their own copy during training and scoring.

### **Team Leader Training and Qualifying**

After the anchor papers, training, and/or qualifying papers were identified and finalized, team leader training began. The scoring director (for each content area or writing type) conducted training for the team leaders. Procedures were similar to those for training scorers (see below) but were more comprehensive, dealing with resolution of discrepant scores, identification of nonscorable responses, unusual prompt treatment, alert situation responses (e.g., child-in-danger), and other duties performed only by team leaders. The team leaders carefully prepared notes on the training papers in preparation for discussion with the scorers, and the scoring director counseled team leaders on training techniques and application of the rubric.

Team leaders assisted in training scorers in team discussions of training sets, and were responsible for distributing, collecting, and accounting for training packets and sample papers during each scoring session. During scoring, team leaders responded to questions, spot-checked reader packets, and counseled scorers having difficulty with the criteria.

Team leaders also administered the quality control (validity sets), monitored the scoring patterns of each reader throughout the project, and conducted retraining as necessary, performed some resolution readings, and maintained a professional working environment. The validity sets were generally selected by the team leaders and scoring director for each content area prior to reader training.

Team leader training lasted from two to four days. Team leaders generally worked 7.75 hours per day, excluding breaks. They set up the room prior to reader arrival each day and meet with scoring directors after scoring each day.

### **Scorer Training and Qualifying**

All scorers were trained using the scoring guides and rubrics, training papers, and/or qualifying papers selected during the rangefinding meetings. Scorers were assigned to a scoring group consisting of one team leader and 10-12 scorers. Each scorer was assigned an individual number for easy identification of their scoring work throughout the scoring session.

After the contracts and nondisclosure forms were signed and the introductory remarks given, training began. Scorer training followed the same format as team leader training except that scorers were not required to annotate each paper in the training sets, although they were encouraged to take notes. The scoring director presented the writing or open-ended item task and introduced the guide, then discussed, room-wide, each score point. This presentation was followed by practice scoring on the training sets. Each scorer was given a clean copy of the scoring guide and training sets, as well as a monitor sheet on which to record training set scores. Because it is easy in a large group to overlook a shy scorer who may be having difficulty, scorers did break into teams to score and discuss the papers in the training sets. This arrangement provided scorers an opportunity to discuss any possible points of confusion or problems in understanding the criteria.

Team leaders collected the monitor sheets after the scoring of each training set and recorded results on a customized log, which was examined by the scoring director to determine which papers were giving scorers difficulty. The scoring director also “floated” from team to team, listening to the team leaders’ explanations and adding additional information when necessary. If a particular paper or type of paper seemed to cause difficulty across teams, the problem was discussed room-wide to ensure that everyone heard the same explanation.

Like team leaders, scorers must demonstrate their ability to score accurately by attaining the agreement percentage established by the New Jersey Department of Education before they may score packets of “live” papers. Any scorer unable to meet these standards was dismissed. All scorers understand this stipulation when they are hired.

Training was carefully orchestrated so that scorers understood how to apply the rubric in scoring the papers, learned how to reference the scoring guide, developed the flexibility needed to deal with a variety of responses, and retained the consistency needed to score all papers accurately.

Scorers were trained to recognize and flag nonscorable responses (fragment, off-topic, not English, no response) and “alert” papers (e.g., suspicion of child abuse) so that these papers could be handled in the correct manner. Alert papers were scored, but then forwarded to the scoring director for review. If the scoring director agreed that the student’s own words specifically stated a situation that qualified as an alert or reflected a potential risk situation for a child, the paper was copied and sent to the Office of Evaluation and Assessment for follow-up with school district personnel. Alert papers are flagged if they reflect potential abuse, emotional or psychological difficulty, dangerous thoughts, or possible plagiarism.

In addition to completing all of the initial training and qualifying, a significant amount of time was allotted for demonstrations of paper flow, explanations of “alerts” and “flagging,” and instructions about other procedures which were necessary for the conduct of a smooth project. Scorer training lasted from two to five days. Scorers generally worked 7.0 hours per day, excluding breaks.

### **Scoring Procedures and Paper Flow**

Each student response was scored by two independent scorers using the scoring scale developed and approved for those items. If the two assigned scores differed by more than one point, the paper was returned for a third “resolution” reading by team leaders or scoring directors. Information about how the three scores were resolved appears in Appendix A.

Before opening a packet, scorers began by writing their assigned reader numbers, as well as the date, on the front of their packet envelope. The stapled packet of papers and the appropriate monitor sheet (first or second reading) was then removed from the envelope. Scorers checked the packet number on the header sheet against the number on the monitor sheet for agreement, and then recorded their scorer identification numbers in the designated space on the scannable monitor sheet. The scorer decided on the score and the assigned scores are recorded in the appropriate spaces provided on the monitor. As scorers progressed through a packet, they checked each paper’s student ID number against the number printed on the monitor sheet. If there was a discrepancy, the packet was flagged for the scoring director to check.

As a scorer completed a packet of papers, he or she returned it to the envelope and gave it to the team leader, along with the monitor sheet. The clerical aide picked up completed packets and monitor sheets and redistribute the packets for second readings.

The packet proceeded to the second reading stage while the first reading scores was being scanned. The procedure for the second reading was the same as that for the first reading, except that the second scorer used the second scoring monitor sheet in the envelope. At no time does the second scorer have access to the scores given by the first scorer. As with the first scoring monitors, the second monitors were scanned and the scores merged into the database.

After the second scores were entered, they were matched with the first scores already in the database. When scores differed by more than one point on any essay, the essay was classified as “discrepant,” a third scoring list by packet and response number was printed, and the response was returned for a third independent reading. After the clerical aide returned the packet to the scoring room, the scoring director located the papers needing a third reading and followed the normal scoring procedures. The third score was scanned in the same manner as the first two scores. The packet was returned to the warehouse and refiled.

### **Scorer Monitoring**

Scorers were monitored in several ways. Team leaders answered scorers' questions, using the guide and training papers as examples. They also read behind their team members by reviewing packets after they were turned in, looking for papers that might merit discussion with the scorer. In addition, every day the scoring director and team leaders received the printout of the scorer statistics—including the scorers' perfect, adjacent and resolution agreement with other scorers, and the scorers' score point distribution. In this way, the scoring director and team leader can look at any one scorer, team, or the room as a whole and rollover items can be compared to previous years.

### **Agreement Between Scorers for the Writing Tasks and Open-Ended Items**

Table 4.2 shows the percentages of writing tasks and open-ended items scored with exact agreement, adjacent agreement, and resolution needed.

The Writing cluster within Language Arts Literacy consists of three writing activities:

- writing/speculate task in response to a picture —  
1 – 6 points, scorer ratings averaged
- writing/persuade task —  
1 – 6 points, scorer ratings summed
- revise/edit —  
0 – 4 points, scorer ratings summed

Each writing task is rated by two independent scorers. Of the more than 350,000 task responses scored for the 2004 administration, 61.8% received exactly the same scores by the scorers and 36.0% received scores that were adjacent. Thus, approximately 97.8% of the task responses required only two scorers. The remaining 2.2% received scores on the Writing Tasks that differed by more than one point and, therefore, required a third scorer.

TABLE 4.2

**Consistency Between Raters Scoring  
GEPA Writing Tasks and Open-Ended Items - 2004**

GEPA Writing Tasks and Open-Ended Items	Percent Raters In Exact Agreement	Percent Raters In Adjacent Agreement	Percent Resolution Needed
<b>Language Arts Literacy</b>			
<b>Writing</b> <b>Total</b>	<b>61.8</b>	<b>36.0</b>	<b>2.2</b>
Writing/Speculate	63.7	35.0	1.3
Writing/Persuade	62.8	35.4	1.8
Revise/Edit	58.8	37.7	3.5
<b>Reading</b> <b>Total</b>	<b>67.1</b>	<b>31.8</b>	<b>1.2</b>
Open-Ended Item 1	66.1	32.5	1.4
Open-Ended Item 2	66.5	31.9	1.7
Open-Ended Item 3	67.5	31.9	0.6
Open-Ended Item 4	68.3	30.8	0.9
<b>Mathematics</b>			
<b>Mathematics</b> <b>Total</b>	<b>88.6</b>	<b>10.3</b>	<b>1.2</b>
Open-Ended Item 11	90.1	9.3	0.7
Open-Ended Item 12	88.4	11.1	0.4
Open-Ended Item 23	89.1	10.1	0.8
Open-Ended Item 24	82.6	16.4	1.1
Open-Ended Item 35	92.9	5.9	1.2
Open-Ended Item 36	88.4	8.9	2.7
<b>Science</b>			
<b>Science</b> <b>Total</b>	<b>79.1</b>	<b>19.5</b>	<b>1.4</b>
Open-Ended Item 1	80.2	18.2	1.6
Open-Ended Item 2	74.2	23.9	2.0
Open-Ended Item 3	82.8	16.5	0.7

All content areas included open-ended items. For the Reading open-ended items, the rubric used by the scorers had score points that ranged from 0 to 4. Two Reading open-ended items are presented for each of two reading passages. For these four items, the resolution percent ranged from 0.6% to 1.7% with the percent at perfect agreement ranging from 66.1% to 68.3%.

Six open-ended items were presented for Mathematics. These six items had percents at perfect agreement ranging from 82.6% to 92.9%. The percent requiring resolution ranged from 0.4% to 2.7%.

Three open-ended items were included for Science. These items had a perfect agreement rate ranging from 74.2% to 82.8%. The percent requiring resolution ranged from 0.7% to 2.0%.

### **4.3 Quality Control Procedures in Data Preparation**

Quality control procedures at Pearson Educational Measurement (PEM) begins with the use of the Capability Maturity Model (CMM), a software development management tool. Key process areas of CMM are requirements management, software project planning, software project tracking and oversight, software quality assurance, and software configuration management. PEM examples of CMM documents include a customer requirements allocation document, a project schedule, functional specifications, a software development project plan, unit test plans, and verification and validation plans. PEM is certified by an external auditor for CMM Level 4, the second highest level of certification.

After software requirements have been identified, the PEM software development team prepares project schedules, project plans, functional specifications, and design documents. PEM begins by creating detailed test plans at both the unit and systems level. A unit test plan is a list of code-unit test cases that is executed and recorded by the software developer. The purpose of the code-unit test process is to ensure that software is developed, maintained, documented, and verified to meet the project requirements for coding and unit testing. As such, the process provides the mechanisms that are necessary to implement the software requirements and design as well as provides code-units quality assurance prior to system test.

After all modules (units) are tested within a system, the CMM process requires a system test. The system test ensures that all the units work together and that outputs from one module match up to the proper inputs for the next module in the system. It also uses expected results to ensure that all requirements have been met. It is important that the system test be performed by a group that is independent of the software development team. This process allows independent verification and interpretation of the requirements. Once the independent testing group has completed the test and given its approval, the system is moved into production mode. It is ready for processing the quality-checking answer documents and files submitted by a quality-checking team.

### **Scanning and Scoring**

Before actual answer documents are machine-scanned, a comprehensive check of the scanning and scoring system is performed. The software development tester creates test decks of gridded answer documents with specific test criteria. The test decks are designed and gridded to cover all response ranges, ID ranges, blanks, and double grids as well as any other responses used by the GEPA. A file containing the scanned responses is then compared to the expected test results for each document to ensure the scanner is operating correctly. The test decks are processed through the programs for scanning and editing answer documents, and packetizing and printing scoring monitors.

The second check involves processing and quality-checking the first actual answer documents received. The NJDOE Office of Evaluation and Assessment and PEM asked 44 districts to return their answer documents early following the test administration so that all test forms could be processed and quality-checked. Also, these early return districts provided the actual student papers for determining score ranges for the writing tasks and open-ended items. Districts were selected to be representative for size and DFG. All information on approximately 60 answer documents are hand checked against the scanned file. In addition, periodically, throughout the processing of the documents, individual answer documents are checked by hand to ensure that scanning is continuing to perform correctly.

### ***NJDOE Quality Control of Score Reporting***

NJDOE Office of Evaluation and Assessment conducted the first round of quality control of multiple-choice items scoring May 3-7, 2004, in New Jersey. PEM printed score sheets for each of the more than 500 students from 16 districts selected by the Office of Evaluation and Assessment for quality control.

Original answer folders for all students in the quality control sample were shipped to the meeting site. PEM maintained a copy of all answer folders in the quality control sample. PEM provided the following materials to the Office of Evaluation and Assessment for the quality control:

1. Scoring masks (punched index and transparency sheets) for all versions of the tests
2. Answer keys for the multiple-choice items
3. Double-grid documentation included a sample of edits for students who marked more than one answer for a multiple-choice item
4. Irregularity reports included all reports dealing with multiple answer folders for students and provided documentation about how these answer folders were merged
5. List of removed items from the Braille and large-print forms
6. List of names of all students taking a Braille or large-print form
7. County-district-school master files with district test coordinators' names and phone numbers
8. Frequency distributions for the student groups, including total, general, LEP, SE, IEP exempts by content area, void counts by reporting category, and Title 1 counts by reporting category.

In the two weeks following the first round of quality control, Measurement Incorporated completed scoring the open-ended and essay responses. Assessment and Evaluation Services equated the test forms after which the NJDOE Office of Evaluation and Assessment and independent reviewers approved the equating procedures and raw score to scale score conversion tables. PEM staff loaded the conversion tables and produced Cycle I score reports for the quality control sample for review.

The second round of the Office of Evaluation and Assessment quality control on the Cycle I score reports took place on May 24 – 28, 2004, at PEM in Iowa City, Iowa. At this time, the open-ended and essay scores were available.

The multiple-choice, open-ended, and essay scores for each cluster and total for the three content areas were systematically checked on all Cycle I score reports. Individual Student Reports for all large-print, Braille, and breach students were produced and reviewed.

Calculations for the Total Scale Score Means and the Just Proficient Means (the mean score for all students across the state whose scale scores were 200 on a particular content area) were verified for each cluster in the content areas by the Office of Evaluation and Assessment staff. Summary statistics included on the School and District Summary Statistics reports were reviewed and approved.

## CHAPTER 5: STANDARD SETTING

### 5.1 Overview of the Process

A proficiency level setting (standard setting) was conducted June 8-11, 1999, to describe and delineate the thresholds of performance that are indicative of Partially Proficient, Proficient, and Advanced Proficient performance for the GEPA Language Arts Literacy and Mathematics. A standard setting study for Science was conducted July 10-12, 2000. Results of these studies were used to formulate recommendations to the Commissioner of Education and the New Jersey State Board of Education for the adoption of the cut scores (i.e., proficiency levels).

The standard setting studies in 1999 and 2000 were conducted by staff from the New Jersey Department of Education, Office of Assessment; Assessment and Evaluation Services; and NCS Pearson. The document, *GEPA Standard Setting Report*, outlines the studies and presents the resulting documentation.

Participants in the standard setting study were chosen because of their qualifications as judges of student performance and content expertise. The judges represented the general population of New Jersey educators. Special care was taken to ensure adequate professional, gender, racial/ethnic, regional, and District Factor Group (DFG) representation on all panels.

A holistic classification method was used for the GEPA standard settings. The judges reviewed student papers sampled to represent the full range of student scores for the March 1999 GEPA administration of the Language Arts Literacy and Mathematics. The judges were asked to classify student work into three categories: Partially Proficient, Proficient, and Advanced Proficient. The judges had the opportunity to review, discuss, and modify their proficiency classifications. Using a logistic regression method, two cut scores were calculated based on judges' classifications. These two cut scores yielded three proficiency levels. Before they finalized their recommended cut scores, the judges examined how their recommended cut scores affected all New Jersey eighth-grade students who took these tests during the first operational administration in 1999.

The methodology and procedures for the Science standard setting study mirrored those used for the Language Arts Literacy and Mathematics standard setting studies. During the Science standard setting in July 2000, judges examined how their recommended cut scores affected all New Jersey eighth-grade students who took the first operational administration of the Science test in 2000.

### 5.2 Procedures

Prior to the standard setting studies, descriptions for Proficient and Advanced Proficient performance were developed by independent panels of eighth-grade language arts, mathematics, and science teachers. The proficiency level descriptors were developed to reflect actual test content. Proficiency level descriptors that are anchored in test content allow for more accurate decisions to be made by the judges. The committees developed the following proficiency level descriptors:

#### ***Language Arts Literacy–Proficiency Level Descriptors***

##### ***Proficient***

Eighth-grade students performing at the proficient level are able to construct meaning as they generate their own texts and work with texts generated by others. Proficient students show an overall understanding of the text at literal and inferential levels. They are able to connect with prior knowledge while interacting with, interpreting, and analyzing text.

In reading exercises, students are able to identify and discuss central themes, supporting details, and organizational structures of text. They can extrapolate and synthesize information, monitor their understanding of text, and identify a purpose for reading. Students at this level are able to identify support for and discuss opinions and conclusions as well as to explain textual conventions and literary elements.

Eighth-grade students proficient in their writing are able to develop a central theme, supporting details, and an organizational structure. They establish and sustain a purpose for writing and elaborate on information as they monitor development of text. Students at this level are able to provide support for opinions and conclusions and to use textual and literary elements appropriately.

##### ***Advanced Proficient***

Eighth-grade students performing at the advanced level are able to construct and extend meaning as they generate their own texts and work with texts generated by others. Advanced students show a sophisticated understanding of abstract themes and ideas that build a text and extend information. They are able to connect with prior knowledge while interacting with, interpreting, analyzing, and critiquing text.

In addition to consistently demonstrating the qualities outlined for a proficient student, the advanced student will demonstrate the ability to synthesize, analyze, and evaluate written text. Students at this level are able to manipulate understanding and will show a high degree of sustained control over textual conventions and literary elements.

**Mathematics–Proficiency Level Descriptors****Proficient**

The student performing at the proficient level demonstrates evidence of conceptual understanding and of procedural and analytic skills. The student demonstrates the ability to apply mathematical skills and knowledge to theoretical and real-world situations. In addition, the student communicates the required skills and makes connections within and among the mathematical content areas.

The student at this level demonstrates a thorough understanding of basic arithmetic operations – an understanding sufficient for problem solving in practical situations. The student understands the connections between fractions, decimals, percents, and other mathematic topics.

The student understands and applies geometric properties and spatial relationships; applies the principles of similarity, symmetry, and coordinate geometry; interprets data and graphs; determines probabilities; applies the concepts and methods of discrete mathematics, and uses algebraic concepts and processes.

**Advanced Proficient**

The student performing at the advanced level demonstrates clear and consistent evidence of thorough conceptual understanding, and of procedural and analytic skills. The student consistently demonstrates the qualities outlined for proficient performance. In addition, the student at the advanced level demonstrates the use of abstract thinking and provides explanations that are consistently clear and thorough.

**Science–Proficiency Level Descriptors**

**Proficient**

The proficient student can recognize the structural levels of living things. This student knows that some traits of organisms are beneficial and some detrimental. This student can interpret visual and textual data to understand the relationship within a food web and the interdependence of living and nonliving systems.

The proficient student can recognize the effect force has on an object, trace the flow of energy through a system, and use the properties of matter to identify and separate materials. This student can understand different types of energy and use information from data charts to interpret relationships and predict outcomes.

The proficient student can recognize the existence of a relationship between the moon and tides, recognize the different characteristics of the planets in the solar system, and understand the natural forces that change the surface of the Earth, including chemical and physical weathering.

**Advanced Proficient**

The advanced proficient student can support scientific conclusions with valid contextual and visual data and make predictions based on the interactions of living things. This student is able to use interpretive skills to analyze visual and textual data in order to solve problems dealing with the application of force and energy.

The advanced proficient student understands the difference between types of energy waves and can recognize and apply experimental principles and empirical data.

The advanced proficient student can recognize the nature of the tides' relationship to Earth, Sun, and moon; interpret topographical maps; and identify the steps in the process of weathering and erosion.

### ***Judge Selection Process and Criteria***

The standard setting process relied on expert judgments. Therefore, nominations were solicited from school districts for teachers or administrators representing excellence in the teaching profession in terms of knowledge of content area, knowledge of eighth-grade students' skills and abilities, and some understanding of assessment procedures. It was considered critical that these judges represent the more general body of expert New Jersey public school educators. Special care was taken to select judges who were representative of the various District Factor Groups (DFGs) within the state. Additionally, districts were specifically asked to include special education, ESL, and bilingual teachers among their nominees. Districts were also encouraged to nominate members of underrepresented minority populations, e.g., African-American or Hispanic, in order to ensure an appropriate diverse representation of statewide populations. Other criteria used in the selection process included number of years teaching experience, the level of content knowledge and student understanding possessed by the nominees, and active participation in content-area professional associations.

Teachers, educators, and content-area experts selected as judges exemplified the required content-area knowledge, teaching experience, and/or understanding of students necessary for an appropriate and comprehensive standard setting study. Each panelist participating in the process represented the knowledge and understanding of his or her peers throughout the course of the process, lending a balance between diverse opinion and consensus.

A concerted effort was made to balance each content-area panel on the basis of county representation, urban representation, representation of schools serving various sizes of populations, gender, and race/ethnicity. The overarching goal of consensus in this forum was not the unanimous agreement of all parties, but the bringing together of individual divergent experiences to form a common understanding of student performance in a content-area that is truly larger, and broader, than its individual parts. The judges selected for the standard setting study represented the same diversity of people and demographics as the students being assessed.

### ***Holistic/Paper Sorting Methodology***

The judges' task was to classify student work into one of three performance categories defined to capture levels of performance as expressed by the Partially Proficient, Proficient, and Advanced Proficient categories. The method was holistic in that the judges considered the whole of an individual student's open-ended and multiple-choice responses, i.e., all the items of a particular student for a content area. With the holistic sorting method, the judges reviewed folders of student papers sampled to represent the full range of scores, and were asked to sort these folders into three performance levels as represented by the quality of the students' work. An outline of the standard setting procedures follows:

### Overview of the 8-Step Plan

#### **Large-Group Session**

The standard setting study began with a large-group session. All judges and participants listened to introductory comments and directions for the three-day meeting. The definitions of the standards, their purpose and ultimate use were discussed. This session was designed to provide a common orientation to judges across content areas.

#### **Step 1 – Description of the Standard Setting Process**

*Judges worked in their own content area and in separate rooms for the remainder of the process. Step 1 provided the judges with an introduction to the process, their role in the process, and a review of the purpose of the standards.*

- *Introductions*
- *Judge Selection Process and Criteria*
- *Purpose of the Standards*
- *Standard Setting Process*
- *Review of the Agenda*
- *Administrative Tasks*

#### **Step 2 – Review of the Assessment Material**

*Judges became familiar with the assessment at this point. They took the assessment under standardized conditions to get a feel for the experience and content. Judges were also introduced to the content validity evidence for the assessment and the open-ended scoring procedures.*

- *Review of Test Content*
- *Brief Description of the Assessment Development Process*
- *Administration of the Assessment to Judges*
- *Scoring the Assessment*

#### **Step 3 – Defining the Standards**

*Step 3 introduced judges to the definitions of the standards. Judges used exercises to brainstorm student work which typified the definitions for each standard. Judges did not write or re-write the definitions at this time. This step only served to familiarize judges with the definitions, which were previously determined, and to help the judges think about students who are at each standard.*

- *Definitions of Student Performance Standards*
- *Interpretation of Proficient Performance*
- *Interpretation of Advanced Proficient Performance*
- *Summary of Student Performance Levels*

#### **Step 4 – Introduction of the Standard Setting Process**

*Step 4 introduced the specific process to the judges. They practiced reviewing student work and sorting student work into three levels of performance – poor, medium, and high. Judges were provided with information about which multiple-choice items were answered correctly on each sample. In addition, scoring rubrics for the open-ended items were reviewed to facilitate the judgment process for the open-ended items.*

- *Description of the Holistic Sorting Method*
- *Summary of the Standard Setting Process*
- *Process Check-off*

**Step 5 – Round 1: Holistic Classification of a Wide Range of Student Papers**

Judges were instructed in the process of completing the rating sheets. Then, judges were given a set of 33 student papers to classify.

The 33 papers were selected to represent the complete range of test scores for each content area. The raw score distribution for a content area was divided into 11 equal intervals. For each interval, three papers were selected to represent a high score, middle score, and low score within the interval. Judges classified each student work sample as representing an Advanced Proficient, Proficient, or Partially Proficient student by the definitions. Judges recorded their classifications on their rating sheets.

Rating sheets were collected and tabulated with results presented to the judges. Classification frequencies for each paper number were shown to the judges. Judges met in small groups to discuss their classifications. Following the discussions, judges were allowed to make changes to their classifications of the student work on their rating sheets.

- *Distribution of Rating Sheets and Instructions*
- *Classification of Papers (Round 1.1)*
- *Discussion of Judges' Ratings*
- *Review of Classifications (Round 1.2)*

**Step 6 – Round 2: Holistic Classification of a Targeted Range of Student Papers**

Based on the judges' ratings from Step 5, preliminary cut scores for Advanced Proficient and Proficient were determined using a logistic response model regression of paper scores upon classification decisions. Two papers from each score point at the preliminary cut score and in a range of 5 score points above and below that cut score were selected. Approximately 22 papers were selected to target the borderline between Advanced Proficient and Proficient and approximately 22 papers were selected to target the borderline between Proficient and Partially Proficient.

Judges were then given the 44 student papers targeted at the preliminary cut scores. Judges classified each of these 44 papers as typical of an Advanced Proficient, Proficient, or Proficient/Partially Proficient student by the definitions. Like Step 5, rating sheets were collected and tabulated with results presented to the judges. Classification frequencies for each paper number were shown to the judges. Judges met in small groups to discuss their classifications. Following the discussions, judges were allowed to make changes to their classifications of the student work on their rating sheets before these were collected.

- *Distribution of Rating Sheets and Instructions*
- *Classification of Papers (Round 2.1)*
- *Discussion of Judges' Ratings*
- *Review of Classifications (Round 2.2)*

**Step 7 – Review of Impact Data**

Judges received reports summarizing their individual ratings and the group cut scores after Step 6. They were provided the statewide performance data to judge the impact of group standards. Judges were allowed, if they desired, to change the raw score value of their cut score according to this new information.

- *Introduction of Individual Judgments and Group Cut Scores*
- *Introduction of Impact Data*
- *Final Standard Determinations*

**Step 8 – Evaluation of the Standard Setting Process**

Judges were encouraged to rate the process using a five-point scale (five being the highest and one being the lowest). Judges were asked to rate the defining and understanding process of Proficient Performance, Advanced Proficient Performance, and Standard Setting Procedures. Finally, they were asked to rate their confidence in the standard setting results. Additionally, open-ended comments were encouraged.

### 5.3 Results

Judges were provided with graphical data depicting the impact of the resulting cut scores on the actual score distributions of New Jersey eighth-grade students. In other words, if the Proficient cut score is X and the Advanced Proficient cut score is Y, then A percent of the students would be Partially Proficient, B percent of the students would be Proficient, and C percent of the students would be Advanced Proficient. The data were based on more than 88,000 students for each of the content areas.

Judges had an opportunity to review the implications of their standards in the form of impact data. Judges received cumulative frequency distributions of student scores that allowed them to see the percent and number of students in each category given the standards the judges had set.

Table 5.1 presents the cut scores determined by the judges at each round of the standard setting. The numbers in the table indicate the Proficient/Advanced Proficient cut scores in raw score points. The judges' ratings were quite stable from Round 1.1 to the final recommended cut score. Table 5.2 shows the percentage of students achieving at each proficiency level for the total population with the final cut scores.

The final cut score recommendations shown in Table 5.1 were approved and adopted by the New Jersey State Board of Education.

**TABLE 5.1**  
**Proficiency-Level Cut Scores**

<b>Cut Scores Proficient/Advanced Proficient</b>	<b>Language Arts Literacy</b>	<b>Mathematics</b>	<b>Science</b>
<b>Total Possible Points</b>	<b>62</b>	<b>56</b>	<b>52</b>
Round 1.1	28.6/45.2	24.4/43.5	24.2/40.1
Round 1.2	28.6/44.7	24.2/43.1	23.7/39.3
Round 2.1	28.2/44.7	24.3/42.8	23.0/39.0
Round 2.2	28.5/45.0	24.5/42.7	24.3/40.2
<b>Final</b>	<b>29.5/44.5</b>	<b>24.0/43.0</b>	<b>24.0/40.0</b>

**TABLE 5.2**  
**Percentage of Students Achieving Each Performance Level**

	<b>Partially Proficient</b>	<b>Proficient</b>	<b>Advanced Proficient</b>
Language Arts Literacy	24.9%	68.8%	6.3%
Mathematics	40.2%	42.7%	17.0%
Science	26.3%	54.5%	19.2%

## CHAPTER 6: SCALING AND EQUATING

### 6.1 Scaling

The individual student scores are reported as scale scores with a range of 100 to 300. The scores 100 and 300 are a theoretical floor and ceiling and may not actually be observed. The scale score of 250 is the cut score between Proficient students and Advanced Proficient students. The scale score of 200 is the cut score between Proficient students and Partially Proficient students. The score ranges are as follows:

<i>Advanced Proficient</i>	<i>250-300</i>
<i>Proficient</i>	<i>200-249</i>
<i>Partially Proficient</i>	<i>100-199</i>

The Partial Credit Model (PCM) is used for scaling and equating the GEPA operational tests. Masters and Wright (1997) provide this description of the Partial Credit Model:

The Partial Credit Model (PCM) is a unidimensional model for the analysis of responses recorded in two or more ordered categories... it belongs to the Rasch family of models and so shares the distinguishing characteristics of that family: separable person and item parameters, sufficient statistics, and, hence, conjoint additivity. These features enable "specifically objective" comparisons of persons and items (Rasch, 1977) and allow each set of model parameters to be conditioned out of the estimation procedure for the other.

The PCM (Masters, 1982, 1987, 1988a, 1988b) is the simplest of all item response models for ordered categories. It contains only *two* sets of parameters: one for persons and one for items. All parameters in the model are *locations* on an underlying variable. (p. 101)

BIGSTEPS was used to provide the Rasch analyses used for generating the item and student statistics.

Raw score to scale score conversion tables for the regular, large-print, Braille, and breach forms are shown in Appendix C. Appendix D shows Language Arts Literacy, Mathematics, and Science scale score frequency distributions.

### 6.2 Equating

Equating designs must take into account the form of the assessment. Two equating designs are used. Mathematics and Science are equated using a common anchor item, non-equivalent group, design in which all students take common items. These common items are selected to be representative of the total test form in terms of content, difficulty, and format.

The structure of Language Arts Literacy does not allow for a subset of common exercises to be selected for use across test administrations because the smallest item exercises are unique and singular. Reading Comprehension is divided into two passage types. These two types cannot be thought of as representative of each other. The Language Arts Literacy equating is accomplished using an embedded equating/field test section that is used for common-item equating.

### Mathematics and Science Equating Design

Common-item equating is used to determine form equivalence from one form, or test administration year, to the next. A set of common (anchor) operational items from the 2003 Mathematics and Science tests was embedded in the 2004 tests. The anchor items include both multiple-choice and open-ended items. Each student participating in the Mathematics and Science testing took the set of common items, and these items contributed to the student's total score. To the maximum extent possible, these items were selected to be proportionally representative of the content and statistics of the total test forms. In addition, the anchor items occupied similar locations in the 2003 and 2004 test forms. These sets of anchor items (14 items with a total of 18 points in Mathematics and 13 items with a total of 15 points in Science) represent approximately one-third of the Mathematics and Science operational tests in terms of number of items and number of points.

The following were applied:

*Calibrate the 2004 test items using the Partial Credit Model and fix the item difficulties to their estimated values based on the 2003 calibration.* The item difficulties for the common anchor items on the 2004 test were fixed to the estimated item difficulties from the calibration of the 2003 operational test. This placed all parameter estimates for the 2004 calibration on the 2003 scale. This also produced the new raw score to ability (theta) table for the 2004 test.

*Develop a raw score to scale score table for the 2004 assessments.* Using the ability to scale score relationship found in the 2003 test calibrations, scale scores were assigned to the raw scores from the 2004 assessments. This was possible because each ability in the ability to scale score table corresponds to a single raw score; therefore, the scale score assigned to that ability can also be assigned to the raw score.

*Common-Item Anchors:* Checks during the equating process were necessary to establish the stability of the common items and determine model fit. One such check was accomplished through the use of the common anchor items from the 2003 operational test embedded in the 2004 operational test. The following is a summary of the steps used for the anchor item analysis.

1. Identify anchor item difficulties from the item bank,
2. Calibrate 2004 form without fixing anchor item difficulties with BIGSTEPS,
3. Calculate mean of the bank anchor items difficulties,
4. Calculate mean of 2004 anchor items,
5. Add constant to 2004 anchor item difficulties so the mean equals that found in the bank values,
6. Subtract 2004 and the bank anchor difficulties after adding the constant,
7. Drop item with largest absolute difference greater than or equal to 0.30 for consideration as anchor item, and
8. Repeat steps 1-7 using remaining anchor items.

The final product from the equating procedure was the raw score to scale score table developed in Step 2. When equating was completed, raw score to scale score conversion tables were available for scoring. These two steps can be applied for future assessments.

### **Language Arts Literacy Equating**

Scaling and equating for Language Arts Literacy was accomplished through a different design. Each assessment has an embedded equating/field test section that is used for either common-item equating or new-item field testing. Language Arts Literacy was equated using a design in which operational items appeared in a section designated for equating or field testing.

The test included the operational items and three equating sections. Students across the state took one of the equating sections or a field test section. Sampling was done by school and stratified by District Factor Grouping to approximate equivalent groups between equating sets. Sample sizes for each equating/field test form were approximately 9,000 students or more than 8 percent of the student examinee population.

The Language Arts Literacy was equated using a common item design with a combined run. Two forms of the 2004 assessment contained two of the operational passages from 2003 in the field test section. Another 2003 field test form contained one of the operational passages for the 2004 administration. This design allowed for the development of a matrix design in the data, with a combination of data records from 2003 and 2004. All data was analyzed in a combined run with the 2003 item parameters fixed to their 2003 values. This places the 2004 item parameters onto the 2003 scale. Using those 2004 item parameters, a raw score to theta relationship was calculated. This was then used to develop the raw score to scale score table.



## CHAPTER 7: TEST STATISTICS

### 7.1 Reliability of the Test Scores

Table 7.1 summarizes reliability estimates for the content areas and clusters. The reliability coefficients given in this table are based on Cronbach's coefficient alpha measure of internal consistency. Cronbach's alpha is used on tests containing items that can be scored along a range of values. The standard errors of measurement (SEMs) for the major content areas are expressed in terms of the raw score metric and the scale score metric. The scale scores range from 100 to 300.

Reliabilities and SEMs for the dichotomously scored items in each cluster are reported using Kuder-Richardson Formula 20 (KR-20) in Table 7.2.

When evaluating these results, it is important to recall that reliability is partially a function of test length. Therefore, the reliability of a content area is likely to be greater than the reliability of a cluster simply because the content area has more items. Similarly, clusters with more items are likely to be more reliable than clusters with fewer items. The data provided in Tables 7.1 and 7.2 reflect the expected positive relationship between test length and reliability.

The SEMs are useful when interpreting students' scores. Measurement error occurs in every test. A student's true score is a hypothetical average score that the student would obtain if a test were repeatedly administered to the student without the effects of instruction, practice, or fatigue. Mehrens and Lehmann (1991) suggest this use of the SEM:

---

The standard error measurement is often used for what is called band interpretation. Band interpretation helps convey the idea of imprecision of measurement...If we assume that the errors are random, an individual's observed scores will be normally distributed about his true score over repeated testing. Thus, one can say that a person's observed scores will lie between  $\pm 1Se$  of his true score approximately 68 percent of the time, or  $\pm 2Se$  of his true score about 95 percent of the time. Of course, we do not know the true score, but one can infer with about 68% (or 95%) certainty that a person's true score is within  $\pm 1Se$  (or  $\pm 2Se$ ) of his observed score. (p. 252)

---

**TABLE 7.1**

**Reliability Estimates and Standard Errors of Measurement (SEMs)  
for Content Areas and Clusters - 2004**

<b>GEPA Test Section</b>	<b>Number of Points</b>	<b>Reliability Cronbach's alpha</b>	<b>SEM Raw Score</b>	<b>SEM Scale Score</b>
<b>Language Arts Literacy</b>	<b>62</b>	<b>.89</b>	<b>3.07</b>	<b>12.29</b>
Reading	36	.87	2.15	-
Writing	26	.79	1.67	-
Interpreting Text	15	.77	1.36	-
Analyzing/Critiquing Text	21	.77	1.66	-
<b>Mathematics</b>	<b>48</b>	<b>.91</b>	<b>3.30</b>	<b>12.80</b>
Number Sense, Concepts, and Applications	12	.72	1.84	-
Spatial Sense and Geometry	12	.72	1.62	-
Data Analysis, Probability, Statistics, and Discrete Mathematics	12	.69	1.60	-
Patterns, Functions, & Algebra	12	.70	1.50	-
Knowledge	48	.91	3.30	-
Problem Solving	38	.89	2.97	-
<b>Science</b>	<b>54</b>	<b>.89</b>	<b>3.29</b>	<b>10.57</b>
Life	19	.78	1.95	-
Physical	19	.72	1.91	-
Earth	16	.71	1.84	-
Cognitive Skills	28	.83	2.43	-
Process Skills	26	.78	2.22	-

TABLE 7.2

**Reliability Estimates and Standard Errors of Measurement (SEMs)  
for Dichotomously Scored Items Within GEPA Content Clusters - 2004**

<b>GEPA Content Area</b>	<b>Number of Items</b>	<b>Reliability (KR-20)</b>	<b>SEM Raw Score</b>
<b>Language Arts Literacy</b>	<b>20</b>	<b>.81</b>	<b>1.73</b>
Reading	20	.81	1.73
Writing*	-	-	-
Writing/Speculate	-	-	-
Writing/Persuade	-	-	-
Revise/Edit	-	-	-
-----			
Interpreting Text	11	.75	1.23
Analyzing/Critiquing Text	9	.58	1.22
<b>Mathematics</b>	<b>30</b>	<b>.87</b>	<b>2.32</b>
Number Sense, Concepts, and Applications	6	.58	1.02
Spatial Sense and Geometry	9	.70	1.33
Data Analysis, Probability, Statistics, and Discrete Mathematics	6	.53	0.98
Patterns, Functions, & Algebra	9	.66	1.26
-----			
Knowledge	30	.87	2.32
Problem Solving	20	.82	1.85
<b>Science</b>	<b>45</b>	<b>.88</b>	<b>3.00</b>
Life	16	.75	1.74
Physical	16	.69	1.82
Earth	13	.67	1.62
-----			
Cognitive Skills	22	.80	2.09
Process Skills	23	.76	2.14

\* There were no dichotomously scored writing items.

## CHAPTER 8: ITEM-LEVEL STATISTICS

The GEPA test specifications are aligned with the Core Curriculum Content Standards. Please refer to the *Technical Manual* and Part 2 of this *Technical Report* for information about the test specifications and test development.

### 8.1 Classical Item Statistics

In Table 8.1, summary statistics are given that describe the difficulty and discrimination of the items comprising each cluster. For dichotomously scored items, means and standard deviations of proportion-correct values (p-values) and point-biserials are given. For the open-ended items, the index of item difficulty is calculated by dividing students' average score on an item by the maximum possible score on the item. Item discrimination for each open-ended item is the correlation between students' item score and their total score on the test section. For both the item-test correlation and the point-biserial correlation, students' total test scores are expressed in terms of the raw score metric.

**TABLE 8.1**

**Item Difficulty and Discrimination Summary Statistics for Dichotomously Scored and Open-Ended Items by Test Section and Cluster - 2004**

GEPA Test Section/Cluster	Dichotomous			Open-Ended		
	Item Difficulty		Item Discrimination	Item Difficulty		Item Discrimination
	Mean	S.D.	Mean	Mean	S.D.	Mean
<b>Language Arts Literacy</b>	<b>.74</b>	<b>.09</b>	<b>.45</b>	<b>.47</b>	<b>.14</b>	–
Reading	.74	.09	.45	.42	.16	.89
Writing	–	–	–	.50	.14	.91
Writing/Speculate	–	–	–	.58	.13	.77
Writing/Persuade	–	–	–	.56	.15	.83
Revise/Edit	–	–	–	.34	.20	.77
Interpreting Text	.76	.09	.49	.43	.16	.71
Analyzing/Critiquing Text	.73	.10	.41	.42	.17	.87
<b>Mathematics</b>	<b>.61</b>	<b>.14</b>	<b>.44</b>	<b>.49</b>	<b>.28</b>	<b>.94</b>
Number Sense, Concepts, and Applications	.65	.14	.45	.46	.38	.81
Spatial Sense and Geometry	.52	.06	.45	.37	.35	.66
Data Analysis, Probability, Statistics, and Discrete Mathematics	.70	.15	.42	.49	.29	.83
Patterns, Functions, & Algebra	.62	.16	.43	.65	.31	.68
Knowledge	.61	.14	.44	.49	.28	.94
Problem Solving	.62	.16	.44	.49	.28	.94
<b>Science</b>	<b>.57</b>	<b>.12</b>	<b>.39</b>	<b>.35</b>	<b>.22</b>	<b>.82</b>
Life	.58	.14	.41	.48	.33	.68
Physical	.57	.10	.37	.21	.22	.54
Earth	.54	.12	.38	.36	.32	.63
Cognitive Skills	.53	.12	.41	.42	.27	.78
Process Skills	.60	.12	.37	.21	.22	.54

Tables 8.2, 8.3 and 8.4 present frequency distributions of item difficulty (p-values) and item discrimination indices by content cluster. The top section of each table shows the distribution of item difficulty values; the bottom section shows the distribution of point-biserial correlations.

Point-biserial indices are produced to evaluate operational test items. Millman and Greene (1989) note that the point-biserial index gives a true reflection of the item’s contribution to the functioning of the test. For field test item review (described in Test Development) biserial correlations are computed. The biserial indices tend to be more stable across samples.

**TABLE 8.2**  
**Frequency Distributions of Item Difficulty and**  
**Item Discrimination by Content Cluster**  
**2004 Language Arts Literacy**

<b>Item Statistics</b>	<b>Interpreting Text</b>	<b>Analyzing Text</b>	<b>Total</b>
<b>ITEM DIFFICULTY: P-VALUES</b>			
.900+	0	0	0
.800 – .899	5	2	7
.700 – .799	3	3	6
.600 – .699	3	3	6
.500 – .599	0	1	1
<.400 – .499	0	0	0
<b>MEAN P-VALUE</b>	<b>.76</b>	<b>.73</b>	<b>.74</b>
<b>MEDIAN P-VALUE</b>	<b>.75</b>	<b>.76</b>	<b>.75</b>
<b>ITEM DISCRIMINATION: POINT-BISERIAL CORRELATIONS</b>			
.50+	3	1	4
.40 – .49	7	3	10
.30 – .39	1	5	6
< .30	0	0	0
<b>MEAN POINT-BISERIAL</b>	<b>.49</b>	<b>.41</b>	<b>.45</b>
<b>MEDIAN POINT-BISERIAL</b>	<b>.48</b>	<b>.39</b>	<b>.47</b>
<b>TOTAL NUMBER OF ITEMS</b>	<b>11</b>	<b>9</b>	<b>20</b>

TABLE 8.3

**Frequency Distributions of Item Difficulty and Item Discrimination by Content Cluster  
2004 Mathematics**

Item Statistics	Number Sense, Concepts, & Applications	Spatial Sense & Geometry	Data Analysis, Probability, Statistics & Discrete Math	Patterns, Functions, & Algebra	Knowledge	Problem Solving	Total Test
<b>ITEM DIFFICULTY: P-VALUES</b>							
.900+	0	0	1	0	1	1	1
.800 – .899	1	0	0	0	1	1	1
.700 – .799	1	0	2	4	7	5	7
.600 – .699	1	1	2	2	6	6	6
.500 – .599	3	5	0	1	9	1	9
.400 – .499	0	3	1	1	5	5	5
.300 – .399	0	0	0	0	0	0	0
< .300	0	0	0	1	1	1	1
<b>MEAN P-VALUE</b>	<b>.65</b>	<b>.52</b>	<b>.70</b>	<b>.62</b>	<b>.61</b>	<b>.62</b>	<b>.61</b>
<b>MEDIAN P-VALUE</b>	<b>.62</b>	<b>.53</b>	<b>.71</b>	<b>.66</b>	<b>.60</b>	<b>.65</b>	<b>.60</b>
<b>ITEM DISCRIMINATION: POINT-BISERIAL CORRELATIONS</b>							
.50 – .59	2	1	1	1	5	3	5
.40 – .49	2	7	3	6	18	13	18
.30 – .39	2	1	1	0	4	2	4
<.30	0	0	1	2	3	2	3
<b>MEAN POINT-BISERIAL</b>	<b>.45</b>	<b>.45</b>	<b>.42</b>	<b>.43</b>	<b>.44</b>	<b>.44</b>	<b>.44</b>
<b>MEDIAN POINT-BISERIAL</b>	<b>.48</b>	<b>.43</b>	<b>.44</b>	<b>.47</b>	<b>.45</b>	<b>.45</b>	<b>.45</b>
<b>TOTAL NUMBER OF ITEMS</b>	<b>6</b>	<b>9</b>	<b>6</b>	<b>9</b>	<b>30</b>	<b>20</b>	<b>30</b>

TABLE 8.4

**Frequency Distributions of Item Difficulty  
and Item Discrimination Indices by Content Cluster**

**2004 Science**

<b>Item Statistics</b>	<b>Life</b>	<b>Physical</b>	<b>Earth</b>	<b>Cognitive Skills</b>	<b>Process Skills</b>	<b>Total Test</b>
<b>ITEM DIFFICULTY: P-VALUES</b>						
.800 +	0	0	0	0	0	0
.700 – .799	3	1	2	1	5	6
.600 – .699	5	6	2	7	6	13
.500 – .599	2	5	2	3	6	9
.400 – .499	4	3	6	8	5	13
<.400	2	1	1	3	1	4
<b>MEAN P-VALUE</b>	<b>.58</b>	<b>.57</b>	<b>.54</b>	<b>.53</b>	<b>.60</b>	<b>.56</b>
<b>MEDIAN P-VALUE</b>	<b>.60</b>	<b>.55</b>	<b>.50</b>	<b>.52</b>	<b>.59</b>	<b>.55</b>
<b>ITEM DISCRIMINATION: POINT-BISERIAL CORRELATIONS</b>						
.60 +	0	0	1	1	0	1
.50 – .59	5	0	1	3	3	6
.40 – .49	3	5	4	6	6	12
.30 – .39	6	8	5	9	10	19
.20 – .29	2	3	1	3	3	6
<.20	0	0	1	0	1	1
<b>MEAN POINT-BISERIAL</b>	<b>.41</b>	<b>.37</b>	<b>.38</b>	<b>.41</b>	<b>.37</b>	<b>.39</b>
<b>MEDIAN POINT-BISERIAL</b>	<b>.40</b>	<b>.39</b>	<b>.37</b>	<b>.39</b>	<b>.35</b>	<b>.39</b>
<b>TOTAL NUMBER OF ITEMS</b>	<b>16</b>	<b>16</b>	<b>13</b>	<b>22</b>	<b>23</b>	<b>45</b>

## 8.2 Speededness

The amount of time allotted for students to complete the test is intended to provide nearly all students with sufficient time to answer all the questions. Table 8.5 presents data concerning the extent to which this intent was met. Open-ended items appear at the end of each part. For this reason, Table 8.5 shows the percentage of students omitting each of the last three multiple-choice items in each part and all open-ended items.

The percent of students omitting the Reading multiple-choice items is very small, at about 0.4%. The percent of students omitting the open-ended items varies from 3.0% to 8.4%.

The percent of students omitting the Mathematics multiple-choice items ranges from 0.2% to 0.5%. The percent of students omitting the Mathematics open-ended items varies from 0.5% to 7.0%.

**TABLE 8.5**  
**Percentage of Students Omitting**  
**the Last Items of Each Test Part - 2004**

Test Section	Multiple-Choice		Open-Ended	
	Item Number	Percentage Omitting	Item Number	Percentage Omitting
<b>Reading</b>				
Part A	Item 8	0.4%	Item 11	3.0%
	Item 9	0.4%	Item 12	8.4%
	Item 10	0.5%		
Part B	Item 8	0.3%	Item 11	3.0%
	Item 9	0.3%	Item 12	6.7%
	Item 10	0.4%		
<b>Mathematics</b>				
Part A	Item 8	0.2%	Item 11	2.3%
	Item 9	0.2%	Item 12	0.5%
	Item 10	0.3%		
Part B	Item 8	0.3%	Item 11	3.5%
	Item 9	0.2%	Item 12	5.4%
	Item 10	0.3%		
Part C	Item 8	0.2%	Item 11	7.0%
	Item 9	0.2%	Item 12	4.1%
	Item 10	0.5%		
<b>Science</b>				
Part A	Item 13	0.5%	Item 16	3.6%
	Item 14	0.7%		
	Item 15	1.1%		
Part B	Item 13	0.3%	Item 16	6.6%
	Item 14	0.5%		
	Item 15	0.5%		
Part C	Item 13	0.3%	Item 16	5.3%
	Item 14	0.5%		
	Item 15	0.6%		

The percent of students omitting the Science multiple-choice items ranges from 0.3% to 1.1%. The percent of students omitting the Science open-ended items varies from 3.6% to 6.6%.

Overall, these data indicate that the amount of time provided for completing the test is appropriate and that speed of response is not a factor that affects students' performances or detracts from the validity of scores.

### 8.3 Intercorrelations

The Pearson product-moment correlation between student scores on Language Arts Literacy and Mathematics was .75, Language Arts Literacy and Science was .71, and Mathematics and Science was .80. Table 8.6 shows the correlations between students' scores in the major content clusters and item types. Table 8.7 shows the correlations between student scores on the content clusters. The scores used for all correlations were expressed in the raw score metric.

Note that correlations between a content area and cluster within that content area are partially a function of the proportion of the content area that is made up of items from the given cluster. Clusters with many items that make up a large proportion of the content area score increase the cluster with content area correlation.

For example, the correlation between Reading and Language Arts Literacy in Table 8.6 is quite high (.97) because 36 Reading points are part of the total Language Arts Literacy 62 points.

In addition, correlations are partially a function of the number of items in the measures being correlated. Therefore, the number of items in the content areas and clusters being correlated must be considered when their correlations are evaluated. In Table 8.7, the L3 Writing/Speculate cluster has only six points, so this cluster may not correlate as highly with other clusters due to this small number of points.

**TABLE 8.6**

#### **Intercorrelations Among Major Content Clusters and Item Types - 2004**

Major Content and Item Types	Major Content Clusters and Item Types										
	Language Arts Literacy					Mathematics			Science		
	LAT	R	R MC	R OE	W	M T	M MC	M OE	ST	S MC	S OE
<b>LAT Language Arts Literacy (62)</b>											
R Reading (36)	.97										
R MC Reading Multiple-Choice (20)	.90	.95									
R OE Reading Open-ended (16)	.89	.88	.69								
W Writing (26)	.91	.77	.67	.77							
<b>MT Mathematics (48)</b>	<b>.75</b>	<b>.72</b>	<b>.68</b>	<b>.65</b>	<b>.66</b>						
M MC Mathematics Multiple-Choice (30)	.71	.69	.66	.60	.62	.96					
M OE Mathematics Open-ended (18)	.72	.69	.63	.65	.64	.94	.82				
<b>ST Science (54)</b>	<b>.71</b>	<b>.71</b>	<b>.69</b>	<b>.61</b>	<b>.64</b>	<b>.80</b>	<b>.77</b>	<b>.74</b>			
S MC Science Multiple-Choice (45)	.69	.69	.67	.59	.61	.78	.76	.72	.99		
S OE Science Open-ended (9)	.65	.64	.59	.56	.59	.68	.65	.66	.82	.73	

Number in parentheses is the number of score points.

Language Arts Literacy N = 108,419; Mathematics N = 108,954; Science N = 108,740.

**TABLE 8.7**  
**Intercorrelations Among Content Areas and Clusters - 2004**

Test Section/Cluster	Test Section/Cluster																				
	Language Arts Literacy							Mathematics							Science						
	LAT	L1	L2	L3	L4	L5	L6	L7	MT	M1	M2	M3	M4	M5	M6	ST	S1	S2	S3	S4	S5
<b>LAT Language Arts Literacy (62)</b>																					
L1 Reading (36)	.97																				
L2 Writing (26)	.91	.77																			
L3 Writing/Speculate (6)	.77	.68	.81																		
L4 Writing/Persuade (12)	.83	.71	.91	.70																	
L5 Revise/Edit (8)	.77	.64	.87	.59	.62																
L6 Interpreting Text (15)	.90	.94	.70	.62	.65	.58															
L7 Analyzing/Critiquing Text (21)	.93	.96	.76	.66	.70	.62	.79														
<b>MT Mathematics (48)</b>	<b>.75</b>	<b>.72</b>	<b>.68</b>	<b>.56</b>	<b>.62</b>	<b>.58</b>	<b>.68</b>	<b>.69</b>													
M1 Number Sense, Concepts, and Applications (12)	.65	.63	.59	.48	.54	.51	.59	.60	.90												
M2 Spatial Sense and Geometry (12)	.61	.59	.55	.45	.50	.48	.56	.56	.88	.70											
M3 Data Analysis, Probability, Statistics, and Discrete Mathematics (12)	.71	.68	.64	.55	.59	.54	.64	.65	.89	.74	.70										
M4 Patterns, Functions, and Algebra (12)	.69	.67	.62	.52	.57	.53	.63	.64	.88	.72	.70	.74									
M5 Knowledge (48)	.75	.72	.68	.56	.62	.58	.68	.69	1.00	.90	.88	.89									
M6 Problem Solving (38)	.75	.72	.68	.56	.62	.58	.68	.69	.99	.89	.85	.90	.87	.99							
<b>ST Science (54)</b>	<b>.72</b>	<b>.71</b>	<b>.61</b>	<b>.52</b>	<b>.55</b>	<b>.53</b>	<b>.67</b>	<b>.68</b>	<b>.80</b>	<b>.70</b>	<b>.72</b>	<b>.71</b>	<b>.70</b>	<b>.80</b>	<b>.79</b>						
S1 Life (19)	.69	.69	.58	.49	.52	.50	.66	.65	.74	.65	.66	.67	.66	.74	.73	.92					
S2 Physical (19)	.63	.62	.55	.46	.49	.48	.59	.59	.72	.63	.65	.63	.63	.72	.71	.90	.73				
S3 Earth (16)	.61	.61	.52	.44	.47	.45	.58	.58	.70	.61	.64	.63	.61	.70	.69	.89	.73	.70			
S4 Cognitive Skills (28)	.68	.68	.58	.49	.52	.50	.64	.64	.77	.67	.70	.68	.67	.77	.76	.96	.89	.84	.87		
S5 Process Skills (26)	.68	.68	.58	.49	.53	.50	.64	.64	.74	.65	.66	.67	.66	.74	.73	.94	.85	.86	.82	.80	

Number in parentheses is the number of score points.  
Language Arts Literacy N = 108,419; Mathematics N = 108,954; Science N = 108,740.

## Chapter 9: Test Validity

The purposes served by the GEPA scores are noted in the following paragraph from page 6 of the manual, *School and District Guidelines: Interpretation and Use of GEPA Results*:

---

The GEPA should serve as a primary indicator for identifying those students who may need instructional intervention. The test should also serve as an indicator for determining which local education programs may need revisions to ensure that instructional programs are aligned with the Core Curriculum Content Standards. The GEPA is intended to evaluate the progress students are making in mastering the knowledge and skills required by the end of the eighth grade and in mastering the knowledge and skills they will need to pass the HSPA.

---

For each of the GEPA content areas, New Jersey educators defined the content and skill test specifications. Content area committees assisted with developing the Directory of Test Specifications and Sample Items which delineate specifications used to create the assessments and to measure student proficiency in the knowledge and skills outlined in the Core Curriculum Content Standards.

Test specifications for the GEPA content areas were designed to align with the Core Curriculum Content Standards. The GEPA Educator Content Committees recommended the emphases and priorities reflected in the number of items for each item type and cluster on the operational test.

The State Board requires that the Core Curriculum Content Standards be reviewed every five years. New standards for the three content areas were adopted by the Board in July 2002. To comply with requirements of the federal No Child Left Behind Act of 2001 (NCLB), the Language Arts Literacy standards were also revised in April 2004. Field test items included for GEPA 2004 were classified originally according to the standards adopted in 1996 and also classified according to the 2002 and 2004 standards.

Curriculum developers and teachers use the specifications, along with Curriculum Frameworks, the standards themselves, and the score reports, to improve instruction at the district, school, and classroom levels. A number of reports have been designed to assist educators with focusing on pertinent information. Report forms designed to meet specific needs extend the effectiveness of a testing program by making it easier to use test results for educational planning. Chapter 10 of this *Technical Report* includes descriptions and examples of the reports.

The GEPA reports include Individual Student Reports, school and district aggregate reports, cluster means reports, and performance reports by deomographic groups. The manual, *School and District Guidelines: Interpretation and Use of GEPA Results*, was developed to assist in the analysis, interpretation, and use of the different types of GEPA score reports. Copies of this manual are included in the shipment of score reports.

Beginning with the 1991 EWT due notice testing, the students' essays also have been returned to the districts for distribution to appropriate district staff members for analysis and use in classroom instruction. A manual, *Cycle II Criterion-Based Holistic Scoring: A Writing Handbook*, included with the essays presents the scoring method and criteria used to evaluate student writing and offers suggestions for using the New Jersey's scoring rubrics and student test data to improve classroom instruction. Teachers are encouraged to review the sample responses in the handbook, the annotations on each of the sample responses, and the features of the respective score scales.

The State Department of Education releases a State Summary Report for each of the tests, which contains district and school results as well as summary results for the state as a whole, District Factor Groups (DFGs), and special needs districts. Districts are required to report test results to their boards of education and to the public within 30 days after receiving test reports. Previously, analysis and interpretation of the school and district reports was required by the New Jersey Administrative Code (N.J.A.C. 6:39-1.4(a)6). Within 45 days of receipt of reports, an analysis had to be completed and a summary report made available to the public.

Further information about the legal and historical background for the GEPA is available at:

<http://www.state.nj.us/njded/code/current/title6a/chap8.pdf>

<http://www.state.nj.us/njded/assessment/history.shtml>

*The Standards for Educational and Psychological Testing* states, "Validity is a unitary concept." (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999, p.11). Since 1991, New Jersey school district personnel received score reports and essays for their eighth grade students from the EWT and GEPA testing programs. These score reports and essays present information for identifying which local education programs are successfully yielding results consistent with the objectives of the New Jersey assessment programs. Score reports and essays assist teachers with information for students' intervention. Information from the item review processes helped item developers and content committee members produce items to measure skills required for eighth grade students in the content areas assessed. The description of the GEPA test specifications development through score reporting suggests there is a firm relationship from GEPA item development through student instruction.

As noted in Chapter 2 of this *Technical Report*, the *Standards for Educational and Psychological Testing* (p. 11-12) recognizes the following possible sources of GEPA validity evidence:

- Evidence based on test content
- Evidence based on response processes
- Evidence based on internal structure
- Evidence based on relations to other variables
- Evidence based on consequences of testing

In the present chapter about validity, discussion of the possible sources of evidence is presented under headings for the traditional validity terms: content and curricular validity, construct validity, criterion-related validity, and consequential validity evidence. The specific sources of GEPA evidence currently included in validity descriptions are identified in the subsequent traditional validity discussions.

### 9.1 Content and Curricular Validity

Content validity is the most relevant and important source of evidence for the GEPA. The validity of the GEPA scores is based on the alignment of the GEPA to the Core Curriculum Content Standards and the knowledge and skills expected of eighth-grade students.

The Core Curriculum Content Standards were developed by teachers and other educational professionals from New Jersey. The Core Curriculum Content Standards outline what students should know and be able to do at a certain grade level. The questions on the GEPA can contain items/concepts included in the grade eight standards as well as for those standards listed for the prior grades.

The content area committees assisted with developing the Directory of Test Specifications and Sample Items for each of the assessed areas. Attributes of New Jersey educators serving on the committees include:

- strong knowledge of the content area,
- familiarity with New Jersey’s Core Curriculum Content Standards for the specific content area,
- understanding of student’s skills and abilities at the eighth-grade benchmark level,
- some understanding of assessment procedures,
- the ability to work effectively in teams,
- a commitment to educational excellence,
- sensitivity to students’ needs.

The three content area directories are available online at:

<http://www.njpep.org/assessment/TestSpecs/LangArts/TOC.html>

<http://www.njpep.org/assessment/TestSpecs/MathTestSpec/GEPA/Math/MathIndex.html>

<http://www.njpep.org/assessment/TestSpecs/ScienceGEPA/index.html>

Sequential procedures of test specification development through operational test approval described in Chapter 2 of this report ensure the content validity of the tests. The item development teams at Measurement Incorporated begin each item development cycle with a review of the Core Curriculum Content Standards and the three directories of test specifications. Using their years of experience with New Jersey item writing and reviews, item writers understand how to develop multiple-choice and open-ended items that tap the appropriate range of skills. They understand the cognitive complexity required within their content area. Items are designed to assess higher-order or critical thinking skills in varied contexts that are familiar to students. Item content for all items, including the writing-task prompts, is carefully reviewed to ensure that items are free from gender, racial, ethnic, and regional bias.

Prior to field testing, all test items are reviewed by the New Jersey Assessment Content and Sensitivity Review Committees as well as the Office of Evaluation and Assessment staff to ensure that items meet GEPA test specifications including appropriate difficulty and skill requirements. Item approval forms used by the Content Review Committees include two categories that address the cognitive complexity of items:

- match to the test specifications
- appropriate difficulty

The Sensitivity Review Committee reviews to ensure that test questions are not offensive and do not reinforce negative stereotypes, and that test questions appropriately reflect multicultural society. Item approval forms used by the Sensitivity Review Committee require each item to be identified as “Definitely Use” or “Revise and Use With Approval” before the item can be included on a field test.

## 9.2 Construct Validity

The glossary of *Standards for Educational and Psychological Testing* (1999) presents this definition of construct validity:

---

A term used to indicate that the test scores are to be interpreted as indicating the test taker's standing on the psychological construct measured by the test. A construct is a theoretical variable inferred from multiple types of evidence, which might include the interrelations of the test scores with other variables, internal test structure, observations of response processes, as well as the content of the test. In the current standards, all test scores are viewed as measures of some construct, so the phrase is redundant with validity. The validity argument establishes the construct validity of a test. (p. 174)

---

A large percentage of the GEPA score points for each content area come from open-ended and essay test questions. Beginning with the rangefinding process and continuing through statistical review, many of the responses to these questions are scored, reviewed, and discussed by the Content Review Committees members, the NJDOE Content Coordinators, and the Measurement Incorporated staff. These processes have been repeated annually since 1993. Information obtained from students' responses to these questions provides insight used for test item acceptance, modification, and rejection as well as for future test item development.

Open-ended questions and essays compose about 68% (42/62) of the Language Arts Literacy points, 38% (18/48) of the Mathematics points, and 17% (9/54) of the Science points. Many open-ended items are field tested each year. During 2004, 32 Reading open-ended items, 23 Mathematics open-ended items, and 11 Science open-ended items were field tested. For each open-ended item, the Measurement Incorporated Project Director prepared a brief summary discussing the types of responses with notes about any issues and concerns. This summary was included with a copy of each item, rubric, sample answer, and rangefinding papers for reference during the statistical review.

For all field test items, Pearson Educational Measurement computed item means, response frequencies, biserial correlations (the field test item with the base test total score), and other descriptive statistics. Content Review Committees used these statistics, their classroom experiences, and the open-ended responses to explain the processes they believed students were using to provide the correct and incorrect responses to items. Committee members reviewed for concerns related to ambiguity, irrelevant clues, and inaccuracy. Each item must be classified as “Definitely Use” or “Revise and Use with Approval” before it could appear on an operational test.

In addition, several statistics including item difficulty, item discrimination, and item omits are produced for the operational test and printed in each *Technical Report*. Other operational statistics calculated include Pearson product-moment correlations between students' scores on the operational test content clusters and item types.

### 9.3 Criterion-Related Validity

The *Standards for Educational and Psychological Testing* (1974) presents this definition of criterion validity:

---

Criterion-related validities apply when one wishes to infer from a test score an individual's most probable standing on some other variable called a criterion. Statements of predictive validity indicate the extent to which an individual's future level on the criterion can be predicted from a knowledge of prior test performance; statements of concurrent validity indicate the extent to which the test may be used to estimate an individual's present standing on the criterion. The distinction is important. (p. 26)

---

Sources of evidence related to concurrent and predictive validity for GEPA score interpretations are linked to the purposes that score report information serves for districts, schools, and teachers. The *Score Interpretation Manual* provides procedures for disseminating score reports and using test score information.

A section using reports for student-level evaluation notes:

---

Further examination of a student's knowledge and skill deficiencies should include the analysis of the student's whole profile. Decisions about appropriate instructional programs should be based on examination of a student's classroom test results, grades, anecdotal records, portfolios, checklists, school-level results, and other measures of performance. (p. 38)

---

One possible source of criterion-related validity is the relationship of GEPA score trends to National Assessment of Educational Progress (NAEP) score trends.

The New Jersey assessments and NAEP have several similarities and major differences. The New Jersey assessments and the NAEP are based on content standards and frameworks that are revised or replaced on a regular basis to keep them in line with current instructional practices. Likewise, both the NAEP and New Jersey assessments create test specifications based on their respective frameworks that provide guidelines for developing the test items.

However, the New Jersey assessments and NAEP are distinctly different assessments because of:

- context and purpose,
- content and skills measured,
- item difficulty and formats,
- method used for setting performance standards (i.e. cut points or achievement levels)

For these reasons, the New Jersey assessments and the NAEP, even in the same content area, may not yield comparable test results.

New Jersey results for the 2003 NAEP Reading and Mathematics tests for grade 8 students included the following:

- **Reading** - The average scale score was 268. About 37 percent of the students scored at or above the NAEP Proficient level while 79 percent of the students scored at or above the NAEP Basic level. Students in two jurisdictions (53 jurisdictions total) attained higher average scale scores than New Jersey students. New Jersey students scored higher than those in 26 jurisdictions and not significantly different from students in 24 jurisdictions.
- **Mathematics** - The average scale score was 281. About 33 percent of the students scored at or above the NAEP Proficient level. In 1990, 21 percent of the students scored at or above the NAEP Proficient level and, in 1992, 24 percent of the students scored at or above the NAEP Proficient level. In 2003, students in eight jurisdictions (53 jurisdictions total) scored higher average scale scores than New Jersey students. New Jersey students scored higher than those in 23 jurisdictions and not significantly different from students in 21 jurisdictions.

Further information about the NAEP and the New Jersey assessments is available online at:

<http://www.state.nj.us/njded/assessment/naep/nj.shtml>

#### 9.4 Consequential Validity Evidence

Messick (1980) noted that test validity is evaluation of evidence and consequence.

Test validity is thus an overall evaluative judgment of the adequacy and appropriateness of inferences drawn from test scores. This evaluation rests on four bases: (1) an inductive summary of convergent and discriminant research evidence that the test scores are interpretable in terms of a particular construct meaning, (2) an appraisal of the value implications of that interpretation, (3) a rationale and evidence for the relevance of the construct and the utility of the scores in particular applications, and (4) an appraisal of the potential social consequences of the proposed use and of the actual consequences when used.

Putting these bases together, we can use test validity to have two interconnected facets linking the source of justification – either evidential or consequential – to the function or outcome of the testing – either interpretation or use. The crossing of basis and function is portrayed in Figure 1.

Figure 1

	Test Interpretation	Test Use
Evidential Basis	Construct Validity	Construct Validity + Relevance/Utility
Consequential Basis	Value Implications	Social Consequences

(p. 1023)

Beginning with the EWT due notice testing in 1991, the EWT and GEPA scores have provided districts information to help align their curriculum and instruction with the content and skills tested. As noted, the manual, *School and District Guidelines: Interpretation and Use of GEPA Results*, was developed to assist in the analysis and interpretation of GEPA score reports. The manual gives examples of uses of test results, discusses the various test scores, provided information about the appropriate score uses, and cautions against inappropriate score use.

Reports such as the District-Designed Reports were developed to provide districts with tools for organizing data to assist with instructional planning. For the 2004 administration, 129 used the “special” code category on the GEPA answer documents to obtain cluster means for selected groups of students. The return of students’ essays for instructional purposes has been an important aspect of Cycle II reporting.

A number of materials including the manual, *Cycle II Criterion-Based Holistic Scoring: A Writing Handbook*, and the Directory of Test Specifications and Sample Items for each of the GEPA content areas give guidance to teachers and curriculum developers for both instructional improvement and alignment.

Longitudinal graphs from 1999 – 2004 for Language Arts Literacy and Mathematics and from 2000 – 2004 for Science are available for the following groups:

- All Students
- Subgroups – General Education, Special Education, Limited English Proficient
- Gender – Female, Male
- Ethnicity – White, Black, Asian, Hispanic
- Economic Status – Economically Disadvantaged, Non-Economically Disadvantaged

The longitudinal graphs for the percent proficient and above by economic status appear in Figure 9.1 for Language Arts Literacy, Figure 9.2 for Mathematics, and Figure 9.3 for Science. The Language Arts Literacy graphs show that the proficient and above scores hovered between 46.2% and 48.1% for the economically disadvantaged students, and between 78.3% and 82.7% for the non-economically disadvantaged students.

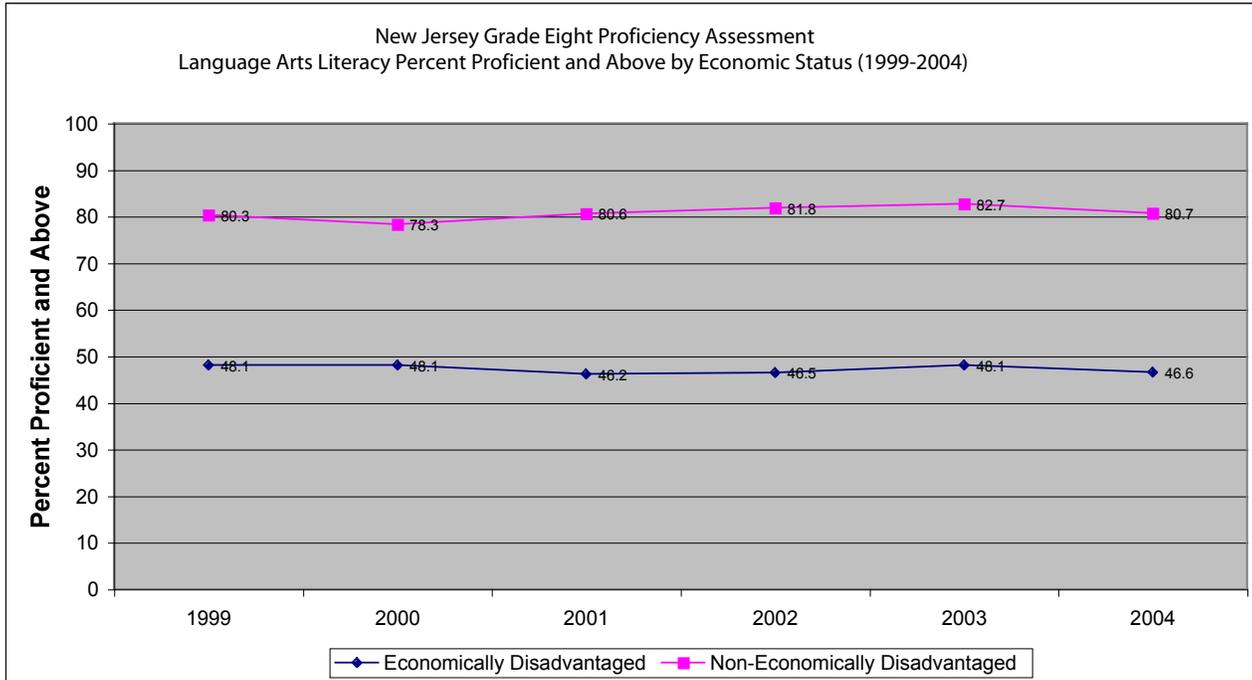
The graphs for Mathematics and Science show generally increasing percents of students with proficient and above scores for the both the economically disadvantaged and non-economically disadvantaged groups. The range of percentages of economically disadvantaged students and non-economically disadvantaged students with proficient and above scores ranged as follows for the 1999-2004 Mathematics administrations and the 2000-2004 Science administrations:

- economically disadvantaged students with proficient and above scores ranged from 25.4% in the 1999 Mathematics test administration to 36.1% in the 2004 test administration;
- non-economically disadvantaged students with proficient and above scores ranged from 64.9% in the 1999 Mathematics test administration to 70.9% in the 2004 test administration;
- economically disadvantaged students with proficient and above scores ranged from 36.9% in the 2000 Science test administration to 50.1% in the 2004 test administration; and
- non-economically disadvantaged students with proficient and above scores ranged from 78.2% in the 2000 Science test administration to 83.6% in the 2004 test administration.

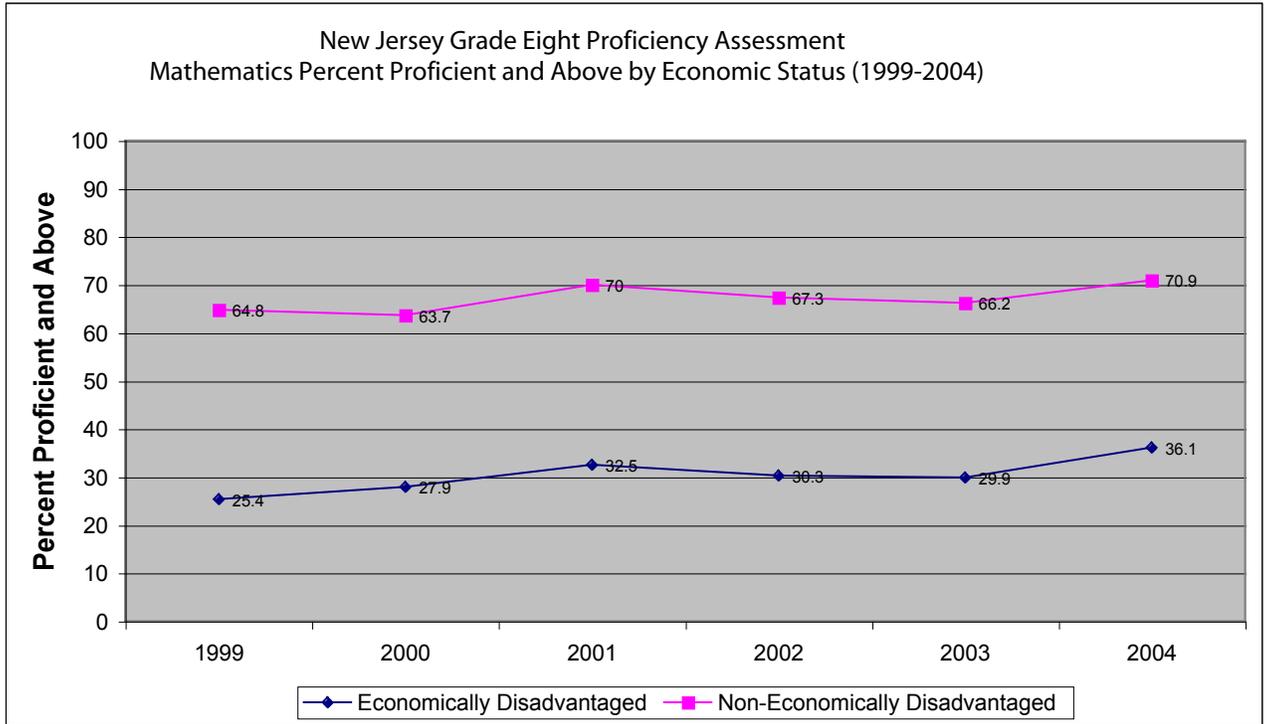
The complete group of longitudinal graphs are available online at:

<http://www.state.nj.us/njded/schools/achievement/2005/gepa/graphs.xls>

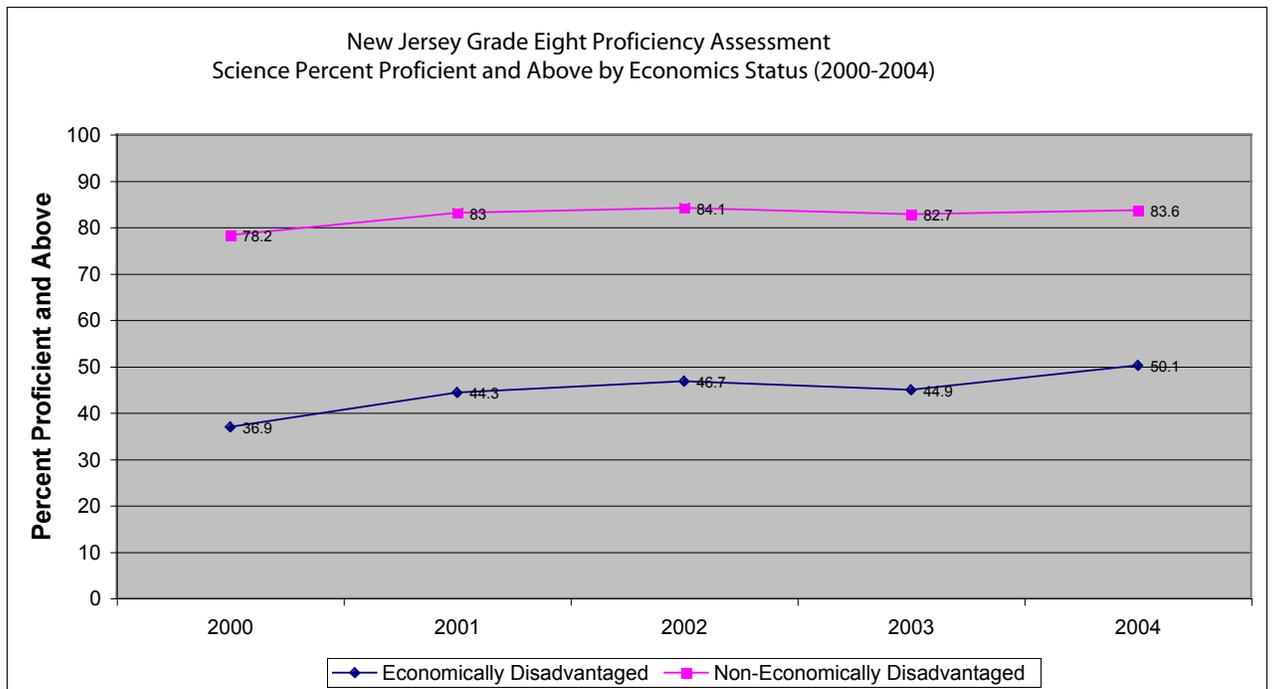
**FIGURE 9.1**  
**LANGUAGE ARTS LITERACY**  
**Longitudinal Graph by Economic Status**



**FIGURE 9.2**  
**MATHEMATICS**  
**Longitudinal Graph by Economic Status**



**FIGURE 9.3**  
**SCIENCE**  
**Longitudinal Graph by Economic Status**



## CHAPTER 10: REPORTING

To help school personnel identify the needs of eighth-grade students tested and to assist in the evaluation of school and district programs, a variety of reports are produced and distributed.

The GEPA reports were produced in two cycles:

- Cycle I reports, including Individual Student Reports and preliminary school and district aggregate reports, were received in the districts in mid-June.
- Cycle II reports, including cluster means reports and performance reports for demographic groups, were received in the districts in late September.

### 10.1 Information on the Reports

The Cycle I and Cycle II score reports are designed to show a range of student identification and score information to assist school personnel with identifying the needs of their students and recognizing weaknesses in instructional programs.

**Student Identification** - Score reports display student identification information gridded on the answer documents or submitted on a pre-ID label files. Prior to reporting, a roster showing the students' demographic information was distributed to school districts to provide an opportunity for corrections.

In addition to the student's name and the Test ID Number assigned to the student, the following information is collected:

- Date of Birth (DOB)
- Gender is indicated by M (male) or F (female).
- Ethnic codes
- <, 1, 2, or 3 (see LEP codes in Appendix E) is indicated in the LEP column if a student was coded as limited English proficient. If multiple bubbles were coded, a Y will appear in this column.
- Y (for yes) is indicated in the TIS<1 column if a student was coded as being enrolled in the school for less than a year.
- Y (for yes) is indicated in the TID<1 column if a student was coded as being enrolled in the district for less than a year.
- A through N (see SE codes in Appendix E) is indicated in the SE column if a student was coded as a special education student.
- The first letter of a content area (L, M, and S) is indicated in the APA/IEP Exempt column if a student was coded as taking the APA in Language Arts Literacy or Math and/or exempt in Science due to an Individualized Education Program (IEP).
- The first letter of a content area (L, M, and S) is indicated in the T-I column if a student was coded as receiving Title I services for any of the three content areas.
- Y (for yes) is indicated in the ED column if a student was coded as Economically Disadvantaged.
- Y (for yes) is indicated for students coded as having Migrant status.
- Y (for yes) is indicated for students coded by their receiving school [public or private] as being an Out of District placement student.
- Y (for yes) is indicated for students coded as being an Out of Residence Placement student.

**Void Codes** – Immediately following testing, examiners mark if a student’s answer document should be voided due to illness, disruptive behavior, or some other reason. The answer folder is not scored and a void code is printed in place of the total test score on the student’s reports. These void codes are as follows:

- V1 (voided due to illness)
- V2 (voided due to cheating or disruptive behavior)
- V3 (voided for some other reason determined by the examiner)
- V5 (voided due to breach of security by a school or district).

Also, a student’s answer document may be voided at the time of scoring. For Mathematics and Science, if a student attempted less than 20 percent of the items, no cluster data will appear and, instead of the content area score, the report will list a V4. For Language Arts Literacy, if a student attempted less than 20 percent of the items on one or two testing days but did attempt 20 percent or more on the other testing day, a V4 will appear instead of the Language Arts Literacy score, but cluster data will be provided on the report.

During the 2004 administration, 445 Mathematics and 102 Science tests were voided due to the attempted criteria. For Language Arts Literacy, 256 tests were voided due to the attempted criteria for Day 1 and 316 tests were voided due to the attempted criteria for Day 2.

**Score Information** – The total GEPA Language Arts Literacy, Mathematics, and Science scores are reported as scale scores with a range of 100 to 300. The scores of 100 and 300 are a theoretical floor and ceiling which may not actually be observed. The scale score of 250 is the cut point between Proficient students and Advanced Proficient students. The scale score of 200 is the cut point between Partially Proficient students and Proficient students. The score ranges are as follows:

---

<i>Advanced Proficient</i>	<i>250 – 300</i>
<i>Proficient</i>	<i>200 – 249</i>
<i>Partially Proficient</i>	<i>100 – 199</i>

---

The scores of students who are included in the Partially Proficient level are considered to be below the state minimum level of proficiency. These students may need additional instructional support, which could be in the form of individual and programmatic intervention. District staff should consider multiple measures for all students before making decisions about students’ instructional placement.

In addition to the total GEPA scores in Language Arts Literacy, Mathematics, and Science, various score reports contain the following information for each cluster (scores at the cluster level are raw scores):

- **Points Earned** – This number represents the number of points a student received for a given cluster. On the Student Roster for Language Arts Literacy, the “Points Earned” is provided for Reading and Writing as well as for each of the writing tasks.
- **Just Proficient Mean** – This number represents the average (mean) number of points received for each cluster by all students in the state whose scale scores are 200 for a particular content. Students who took Large-Print or Braille forms are excluded from calculating just proficient means.

**Automatic Rescores** – Beginning in 2003, GEPA adopted automatic rescoring of all open-ended responses for all students who receive a scale score ranging from 197 to 199. This automatic scoring process provides an opportunity to detect possible scoring anomalies and afford every eligible student the benefit of another examination and additional consideration of their open-ended responses.

## 10.2 Types of Reports

### Cycle I Reports

#### Individual Student Report (ISR) and Student Sticker

The Individual Student Report (ISR) is a two-sided report showing specific student score information on the front of the ISR. A description of the GEPA and an interpretation of the ISR scores are printed on the back. Figure 10.1 presents the front of a student’s sample report with demographic information, scale scores, proficiency levels, and cluster raw scores and Just Proficient Means. Figure 10.2 shows the GEPA description and ISR interpretation printed for all students.

Two copies of the ISR are produced for every student tested. After educators and school staff analyze the score information on the front of the ISR, one copy is placed in the student’s permanent folder and the other copy is shared with the student’s parent/guardian in a manner determined by the local district. When a student attends a private school as an Out of District Placement student, a third copy of the ISR is produced and sent to the private school.

A student’s scale scores and proficiency levels with the student’s identification information are printed on a peel-off label for attaching to a student’s permanent folder.

#### All Sections Roster

The All Sections Roster, an alphabetical listing of students’ names, provides students’ identification and score information. Each student’s scale scores with proficiency levels are listed for the three content areas. Users of this report can quickly determine how a particular student performed in each of the three content areas. The All Sections Roster provides the most complete listing of the student identification information with codes.

**FIGURE 10.1**  
**Individual Student Report (ISR Front)**



**New Jersey Statewide Assessment System**  
**Grade Eight Proficiency Assessment**  
**Individual Student Report**

**Test Date:** March 2004  
**County:** 99 MIDSTATES  
**District:** 9999 MIDSTATES  
**School:** 999 MIDSTATES M.S.  
**Student Name:** BRADLEY, DENIECE  
**Date of Birth:** 01/11/YY  
**Gender:** F  
**LEP:**  
**SE:**  
**APA/IEP Exempt:**  
**Title I:**  
**Test ID No.:** 0007381924  
**Answer Folder No.:** 054321  
**District/School Student ID No.:**

Content Area	Your Score	Proficiency Level
Language Arts Literacy	235	Proficient
Mathematics	243	Proficient
Science	242	Proficient

**Partially Proficient: Score BELOW 200**  
**Proficient: Score AT OR ABOVE 200**  
**but BELOW 250**  
**Advanced Proficient: Score AT OR ABOVE 250**

Language Arts Literacy	Mathematics	Science																																																						
<p>The Language Arts Literacy section assesses a student's abilities in the following clusters. A check mark indicates the areas of possible strength.</p> <table border="1"> <thead> <tr> <th>Cluster</th> <th>Your Score</th> <th>Just Proficient Mean</th> </tr> </thead> <tbody> <tr> <td>Writing</td> <td>15.0 out of 26</td> <td>12.2</td> </tr> <tr> <td>Reading</td> <td>25.0 out of 36</td> <td>14.3</td> </tr> <tr> <td>Interpreting Text</td> <td>10.0 out of 14</td> <td>5.9</td> </tr> <tr> <td>Analyzing/Critiquing Text</td> <td>15.0 out of 22</td> <td>8.4</td> </tr> </tbody> </table>	Cluster	Your Score	Just Proficient Mean	Writing	15.0 out of 26	12.2	Reading	25.0 out of 36	14.3	Interpreting Text	10.0 out of 14	5.9	Analyzing/Critiquing Text	15.0 out of 22	8.4	<p>The Mathematics section assesses a student's abilities in the following clusters. A check mark indicates the areas of possible strength.</p> <table border="1"> <thead> <tr> <th>Cluster</th> <th>Your Score</th> <th>Just Proficient Mean</th> </tr> </thead> <tbody> <tr> <td>Number Sense, Concepts, &amp; Applications</td> <td>11.0 out of 12</td> <td>6.3</td> </tr> <tr> <td>Spatial Sense &amp; Geometry</td> <td>8.5 out of 12</td> <td>4.7</td> </tr> <tr> <td>Data Analysis, Probability, Statistics, &amp; Discrete Mathematics</td> <td>9.0 out of 12</td> <td>7.4</td> </tr> <tr> <td>Patterns, Functions, &amp; Algebra</td> <td>8.5 out of 12</td> <td>5.6</td> </tr> <tr> <td>Knowledge</td> <td>37.0 out of 48</td> <td>24.0</td> </tr> <tr> <td>Problem Solving Skills</td> <td>18.5 out of 22</td> <td>10.3</td> </tr> </tbody> </table>	Cluster	Your Score	Just Proficient Mean	Number Sense, Concepts, & Applications	11.0 out of 12	6.3	Spatial Sense & Geometry	8.5 out of 12	4.7	Data Analysis, Probability, Statistics, & Discrete Mathematics	9.0 out of 12	7.4	Patterns, Functions, & Algebra	8.5 out of 12	5.6	Knowledge	37.0 out of 48	24.0	Problem Solving Skills	18.5 out of 22	10.3	<p>The Science section assesses a student's abilities in the following clusters. A check mark indicates the areas of possible strength.</p> <table border="1"> <thead> <tr> <th>Cluster</th> <th>Your Score</th> <th>Just Proficient Mean</th> </tr> </thead> <tbody> <tr> <td>Life Science</td> <td>16.0 out of 19</td> <td>8.6</td> </tr> <tr> <td>Physical Science</td> <td>15.0 out of 19</td> <td>7.4</td> </tr> <tr> <td>Earth Science</td> <td>7.0 out of 16</td> <td>6.0</td> </tr> <tr> <td>Cognitive Skills</td> <td>19.0 out of 30</td> <td>11.3</td> </tr> <tr> <td>Process Skills</td> <td>19.0 out of 24</td> <td>10.7</td> </tr> </tbody> </table>	Cluster	Your Score	Just Proficient Mean	Life Science	16.0 out of 19	8.6	Physical Science	15.0 out of 19	7.4	Earth Science	7.0 out of 16	6.0	Cognitive Skills	19.0 out of 30	11.3	Process Skills	19.0 out of 24	10.7
Cluster	Your Score	Just Proficient Mean																																																						
Writing	15.0 out of 26	12.2																																																						
Reading	25.0 out of 36	14.3																																																						
Interpreting Text	10.0 out of 14	5.9																																																						
Analyzing/Critiquing Text	15.0 out of 22	8.4																																																						
Cluster	Your Score	Just Proficient Mean																																																						
Number Sense, Concepts, & Applications	11.0 out of 12	6.3																																																						
Spatial Sense & Geometry	8.5 out of 12	4.7																																																						
Data Analysis, Probability, Statistics, & Discrete Mathematics	9.0 out of 12	7.4																																																						
Patterns, Functions, & Algebra	8.5 out of 12	5.6																																																						
Knowledge	37.0 out of 48	24.0																																																						
Problem Solving Skills	18.5 out of 22	10.3																																																						
Cluster	Your Score	Just Proficient Mean																																																						
Life Science	16.0 out of 19	8.6																																																						
Physical Science	15.0 out of 19	7.4																																																						
Earth Science	7.0 out of 16	6.0																																																						
Cognitive Skills	19.0 out of 30	11.3																																																						
Process Skills	19.0 out of 24	10.7																																																						

Note: The scores in this report are for illustrative purposes only. For example, the raw score to scale score conversions are not necessarily the same as those used in the actual reports

## FIGURE 10.2 Individual Student Report (ISR Back)

### ABOUT THE GRADE EIGHT PROFICIENCY ASSESSMENT

The Grade Eight Proficiency Assessment (GEPA) was administered in March 2004 to approximately 108,000 eighth graders throughout New Jersey. The 2004 GEPA measured students' knowledge and skills in three important content areas --- Language Arts Literacy, Mathematics, and Science. The GEPA focuses on the knowledge and skills needed for future success in high school and beyond high school.

The GEPA Language Arts Literacy section measures both reading and writing. The Reading component asks students to read real-world texts and to answer related questions. It contains both multiple-choice and open-ended items. The open-ended items require students to write a few sentences or a few paragraphs to answer a question about the text. The Writing component asks students to write two essays as well as edit the writing of another author. All the tasks in the Writing component require students to write a response.

The GEPA Mathematics section measures students' abilities to solve problems using mathematical concepts. The components in this section measure: Number Sense, Concepts, and Applications; Spatial Sense and Geometry; Data Analysis, Probability, Statistics, and Discrete Mathematics; and Patterns, Functions, and Algebra. The Mathematics section, like the Reading component of the Language Arts Literacy section, contains both multiple-choice and open-ended items. The open-ended items require students to solve a problem as well as explain their solution.

The GEPA Science section measures students' knowledge and skills in Life Science, Physical Science, and Earth Science. The Science section contains both multiple-choice and open-ended items. The open-ended items require students to respond to a question as well as explain their answer.

Carefully trained readers score the open-ended items and essays. Two readers independently score each student's response to ensure the accuracy of each student's results.

The GEPA serves as an early warning for possible difficulties on the High School Proficiency Assessment (HSPA). Beginning with students who entered their junior year in 2001-2002, the HSPA is the "test of record" for receiving a high school diploma. All students are required to receive a score of "proficient" or higher on each section of the HSPA to earn a New Jersey-endorsed high school diploma. (The HSPA replaces the current High School Proficiency Test or HSPT.)

### HOW TO READ THIS REPORT

This *Individual Student Report* presents your child's Language Arts Literacy, Mathematics, and Science scores on the 2004 GEPA. The report is available only to parents, guardians, students, and authorized school personnel. If you have any questions about the report or how to interpret the scores, you should contact your child's teacher or principal.

Your child's total GEPA Language Arts Literacy, Mathematics, and Science scores are presented in the box on the top half of the report. In the column labeled **Your Score**, your child's scale score for each GEPA section is printed. To the right of the scale score is a column labeled **Proficiency Level**. If the scale score is below 200, your child is "Partially Proficient" in that content area. If the scale score is at or above 200 but below 250, your child is "Proficient" in that content area. Finally, if the scale score is at or above 250, your child is "Advanced Proficient" in that content area. Scores below 200 may indicate a need for additional instructional assistance. However, like any single test score, GEPA results should not be used as the sole basis for instructional decisions.

Additional information to assist in identifying your child's strengths and weaknesses is presented on the bottom half of the report. Cluster-level results show how your child performed on sets of items that measure particular knowledge and skills (clusters above the line) or particular processes (clusters below the line). Though an item on the GEPA can contribute to a cluster above the line (for example, Reading) as well as a cluster below the line (for example, Interpreting Text), each item is counted only once in the total score.

For each cluster, the column to the right of the cluster name, labeled **Your Score**, shows the number of points your child earned on the clusters. The right-most column for each content area, labeled **Just Proficient Mean**, is a yardstick against which to measure your child's cluster score. It is calculated by taking the average of the raw scores of all the students across the state whose scale scores are 200 for a particular content area. For example, among all students who scored at 200 for the GEPA Language Arts Literacy section, the reading cluster score was 19.5. If your child scored at or above 19.5 for this cluster, a check mark (✓) indicates this area is a possible strength for your child. If your child scored below 19.5 on this cluster, he or she may need additional help in this area.

At the top of the report, a notation may appear if, for some reason, your child's test booklet was not scored. These reasons include illness (V1), disruptive behavior (V2), or some other reason (V3). For Mathematics and Science, if a student attempted less than 20% of the items, no cluster data will appear and instead of **Your Score** the report will indicate a V4. For Language Arts Literacy, if a student attempted less than 20% of the items on each of the two testing days, no cluster data will appear and instead of **Your Score** for Language Arts Literacy the report will indicate a V4. If a student attempted at least 20% of the items on one of the two testing days, but did not attempt 20% on the other testing day, a V4 will appear in **Your Score** for Language Arts Literacy but cluster scores will be provided on the report. If a student did not take a section of the test, no cluster data will appear and instead of your score, the report will indicate **Not Present**.

### **Student Roster – Language Arts Literacy, Mathematics, and Science**

Three Student Rosters are printed – one for each content area. Students' names are listed in descending order of the content area scores. Figure 10.3 shows an example of the Student Roster – Mathematics listing the student with the highest score mathematics score first followed with the other students in this school. A dashed line is printed across the roster after the last student in each proficiency level.

No students in the example shown in Figure 10.3 had scores at or above 250, the Advanced Proficient cut point, so a dashed line is printed across the top of the roster. Another dashed line appears across the roster under 200, the Proficient cut point. Students whose answer documents were voided and students who were coded indicating they were taking the Alternate Proficiency Assessment (APA) are listed alphabetically at the end of each content area roster.

### **Summary of School Performance and Summary of District Performance**

A Summary of School Performance is printed for each of the three content areas and a Summary of District Performance is printed for each of the three content areas. The report for each content area provides the number and percent of students in each proficiency level as well as the number of general education students, special education students, and limited English proficient students tested for the content area.

The total test information includes the school or district mean for the reported content area. In addition, the means are provided for each of the clusters. The total test and cluster means are printed for the four student groups: total, general education, special education, and limited English proficient.

The following summary information is provided for each subgroup shown on the report:

- Number Enrolled: total number of answer folders returned
- Number Not Present: number of answer folders returned that were totally blank excluding answer folders coded as APA/IEP exempt
- Number of Voids: number of answer folders coded void by the school [V1, V2, and V3] AND coded void due to less than 20% of the test items being taken, including answer folders coded as APA/IEP Exempt [V4] AND coded void due to a security breach [V5]. Number of Valid Score Scores: total number of students tested excluding not present and voids
- Total number of students who scored in each proficiency level, excluding students coded as APA/IEP Exempt
- Percent of students who scored in each proficiency level, excluding students coded as APA/IEP Exempt

### **Preliminary Performance by Demographic Groups – State Report**

This preliminary report is produced with the Cycle I reports prior to the completion of the automatic rescoring. The one-page report presents the results for the total, general education, special education, and limited English proficient student groups, and by gender, migrant status, ethnicity, and economic status. Data are based on scale scores and the percentage of students that fall into each of the three proficiency levels. The report does not show cluster level data.

**FIGURE 10.3**  
**Student Roster (Mathematics)**

**New Jersey Statewide Testing System  
Grade Eight Proficiency Assessment  
Student Roster – Mathematics**



TEST DATE: MARCH 2004  
REPORT PRINTED:  
COUNTY: 99 MIDSTATE  
DISTRICT: 9999 MIDSTATE  
SCHOOL: 9999 MIDSTATE E.S.

PAGE: 1

STUDENTS PROCESSED: 430

POINTS EARNED

STUDENT NAME/ TEST ID NO.	GENDER	DATE OF BIRTH	LEP	SE	APA	T-I	POINTS EARNED						TOTAL SCORE				
							NUMBER SENSE, CONCEPTS & APPLICATIONS	SPATIAL SENSE & GEOMETRY	DATA ANALYSIS, PROBABILITY, & DISCRETE MATHEMATICS	PATTERNS, FUNCTIONS & ALGEBRA	KNOWLEDGE	PROBLEM SOLVING					
GORELOV 0005739748	I	06/30/YY					6.0	7.0	12.0	8.5	35.5	14.0	12 <sup>1</sup> 4.7 <sup>2</sup>	12 <sup>1</sup> 5.6 <sup>2</sup>	48 <sup>1</sup> 24.0 <sup>2</sup>	38 <sup>1</sup> 18.3 <sup>2</sup>	232
TIPPER 0005739748	F	06/30/YY					6.0	7.0	12.0	8.5	35.5	14.0	12 <sup>1</sup> 4.7 <sup>2</sup>	12 <sup>1</sup> 5.6 <sup>2</sup>	48 <sup>1</sup> 24.0 <sup>2</sup>	38 <sup>1</sup> 18.3 <sup>2</sup>	218
DICLEMENTE 0005739705	N	03/25/YY					11.0	4.0	6.0	8.0	29.0	16.0					213
ELVIS 0005739713	M	10/30/YY					7.0	3.0	10.0	7.5	27.5	10.5					208
GAIL 0005739519	E	09/28/YY					4.0	6.5	10.0	5.0	25.5	9.5					206
MCGLYNN 0005739519	F	09/28/YY					6.0	5.0	10.0	4.0	25.0	11.0					192
SPINELLI 0005739845	F	02/20/YY					5.0	5.0	7.0	3.0	20.0	8.0					186
SCOTT 0005991927	L	05/16/YY					7.0	2.0	6.0	3.0	18.0	7.0					183
RALPH 0005991927	M	05/16/YY					6.0	1.0	6.0	4.0	17.0	7.0					172
PERRY 0005991269	M	08/20/YY					2.0	3.0	4.0	4.0	13.0	7.0					169
KEITH 0005992893	M	03/23/YY					1.0	2.5	4.5	4.0	12.0	6.5					168
BARNES 0005992893	M	03/23/YY					3.0	2.0	3.5	3.0	11.5	3.0					155
BURKE 0005992923	A	12/25/YY					3.0	0.0	2.0	2.0	7.0	3.0					V4
DURBUN 0005991838	N	05/21/YY					3.0	0.0	2.0	2.0	7.0	3.0					V4
CASSIE 0005991838	N	05/21/YY					3.0	0.0	2.0	2.0	7.0	3.0					V4
ROSS 0005991552	S	08/31/YY					3.0	0.0	2.0	2.0	7.0	3.0					V4
BETSY 0005991552	S	08/31/YY					3.0	0.0	2.0	2.0	7.0	3.0					V4
TAYLOR 0005991561	R	05/14/YY		I			3.0	0.0	2.0	2.0	7.0	3.0					V4
JAMES 0005991561	R	05/14/YY					3.0	0.0	2.0	2.0	7.0	3.0					V4
ALSTON 0005992885	R	03/02/YY					3.0	0.0	2.0	2.0	7.0	3.0					V4
ALVIN 0005992885	R	03/02/YY					3.0	0.0	2.0	2.0	7.0	3.0					V4

<sup>1</sup> THE NUMBERS IN THIS ROW ARE THE NUMBER OF POINTS POSSIBLE.

<sup>2</sup> THE NUMBERS IN THIS ROW ARE THE STATEWIDE RAW SCORE MEANS FOR STUDENTS WHOSE SCALE SCORE IS 200.

Note: The scores in this report are for illustrative purposes only. For example, the raw score to scale score conversions are not necessarily the same as those used in the actual reports.

### **Preliminary Performance by Demographic Groups – School and District Reports**

This report is produced before the rescore is completed. This report does not break the data out at the cluster level. Data are based on scale scores and the percentage of students who fall into each of the three proficiency levels.

### **Cycle II Reports**

The Cycle II reports include a final Performance by Demographic Groups report that reflects any changes that may have occurred during the processing of automatic rescoring.

### **School and District Cluster Means Reports**

Figure 10.4 shows an example of the School Cluster Means Report – Language Arts Literacy. The School and District Cluster Means reports consist of three reports – one for each content area.

The first column on the report presents the mean cluster scores for students in the state whose scale score is 200, i.e., students who are “just proficient.” Data include raw score means of all students (total, general education, special education, and limited English proficient student groups) at the cluster level for each content area. A similar format is used for both the School Reports and District Reports. The District Reports present aggregated data for the district, DFG, and the state. Additionally, the School Reports show school level data.

### **District-Designed Reports**

The District-Designed Reports are similar to the School Cluster Means Reports except schools create the reports for selected groups of students. Schools used a “special” code category on the GEPA answer documents to obtain cluster means for selected student groups. Like the School Cluster Means Reports, a District-Designed Report is produced for each content area.

Student answer documents may be coded in any of the four two-column “Special Codes” grids labeled A, B, C, or D. These special codes were assigned by the school during the test administrations. The special code, as coded on the students’ answer folders, is printed in the report title. Student groups must contain six or more students.

**FIGURE 10.4**  
**Cluster Means Report**

**New Jersey Statewide Testing System  
Grade Eight Proficiency Assessment  
Cluster Means<sup>1</sup>**



Test Date: March 2004

CYCLE II

DISTRICT: 99-Midstate

SCHOOL: 999-Midstate Middle School

Number Enrolled: 429

School

Clusters	JUST PROFICIENT MEAN <sup>2</sup>	SCHOOL MEAN	DISTRICT MEAN	DFG MEAN	SPECIAL NEEDS MEAN	NON SPECIAL NEEDS MEAN	STATE MEAN
<b>Language Arts Literacy</b>							
1. Writing (26*)	12.2						
Total Students <sup>3</sup>		9.9	9.9	11.2			13.5
General Education		10.5	10.5	12.2			14.9
Special Education		7.3	7.3	7.2			9.4
Limited English Proficient		8.8	8.8	9.2			9.5
Title I							
2. Reading (36*)	14.3						
Total Students <sup>3</sup>		11.9	11.9	14.9			19.5
General Education		12.8	12.8	14.9			11.8
Special Education		8.1	8.1	8.6			11.2
Limited English Proficient		9.0	9.0	10.3			18.3
Title I							
3. Interpreting Text (15*)	5.9						
Total Students <sup>3</sup>		5.2	5.2	5.6			7.6
General Education		5.6	5.6	5.2			8.1
Special Education		3.6	3.6	3.7			5.0
Limited English Proficient		4.1	4.1	4.3			4.7
Title I							
4. Analyzing/Critiquing Text (21*)	8.4						
Total Students <sup>3</sup>		6.7	6.7	8.7			11.6
General Education		7.2	7.2	8.7			5.9
Special Education		4.4	4.4	5.0			6.5
Limited English Proficient		4.9	4.9	6.0			10.8
Title I							

<sup>1</sup> Refers to the total raw score points in each cluster.

<sup>2</sup> Cluster means exclude students who took Large print, Braille, and alternate forms as well as students coded as voids.

<sup>3</sup> Just Proficient Means are statewide raw score means for students whose scale score is 200.

<sup>4</sup> Students appear in each applicable category, but they are included in Total Students only once.

### **Performance by Demographic Groups – State Report**

Performance by Demographic Groups – State Report summarizes statewide total population data collected from districts regarding general education (GE), special education (SE), LEP, gender, migrant status, ethnicity, and economic status (disadvantaged vs. not disadvantaged). This report includes data from all three content areas. Data are based on scale scores and the percentage of students who fall into each of the three proficiency levels. The report does not break out the data at the cluster level.

The Cycle II Test Results in Appendix B include the Performance by Demographic Groups – State Report.

### **Performance by Demographic Groups – School and District Reports**

Performance by Demographic Groups – School and District Reports present results by general education, special education, LEP, gender, migrant status, ethnicity, and economic status (disadvantaged vs. not disadvantaged) for all three content areas. These group reports provide additional achievement information that can be used to make adjustments to curricula that may better serve these subsections of the total student population. Figure 10.5 shows an example of the Performance by Demographic Groups – School Report.

Similar to the Performance by Demographic Groups – State Report, data included are based on scale scores and the percentage of students who fall into each of the three proficiency levels. The reports do not break out the data at the cluster level.

**FIGURE 10.5**  
**Performance by Demographic Groups**

TEST DATE: MARCH 2004

CYCLE II

REPORT PRINTED

COUNTY: 99 MIDSTATE

DISTRICT: 9999 MIDSTATE

SCHOOL: 999 MIDSTATE M.S.

NUMBER ENROLLED : 379

**New Jersey Statewide Testing System  
Grade Eight Proficiency Assessment  
Performance by Demographic Groups  
School**



	Language Arts Literacy					Mathematics					Science				
	Number Enrolled	Number Not Present	Number of Voids	Number of Valid Scale Scores	% Proficient	% Partially Proficient	% Advanced Proficient	Scale Score Mean	Number Not Present	Number of Voids	Number of Valid Scale Scores	% Proficient	% Partially Proficient	% Advanced Proficient	Scale Score Mean
<b>Total Students<sup>1</sup></b>	366	0	1	365	73.0	73.0	0.0	185.8	1	0	365	85.5	85.5	1.6	177.1
General Education	359	0	1	358	73.0	73.0	0.0	185.8	1	0	358	73.0	73.0	0.0	185.8
Special Education	4	0	0	4	50.0	50.0	0.0	143.0	0	0	4	50.0	50.0	0.0	143.0
Limited English Proficient <sup>2</sup>	3	0	0	3	100.0	100.0	0.0	180.2	0	0	3	100.0	100.0	0.0	180.2
<b>Gender<sup>3</sup></b>															
Female	185	0	1	184	69.7	69.7	0.0	188.5	1	0	184	87.3	87.3	1.1	176.2
Male	181	0	0	181	76.2	76.2	0.0	183.1	0	0	181	83.6	83.6	2.1	178.0
<b>Migrant Status</b>															
Migrant	0	0	0	0	0.0	0.0	0.0	0.0	0	0	0	0.0	0.0	0.0	0.0
Non-Migrant	366	0	1	365	73.0	73.0	0.0	185.8	1	0	365	85.5	85.5	1.6	177.1
<b>Ethnicity</b>															
White	8	0	1	7	50.0	50.0	0.0	193.8	1	0	7	66.7	66.7	0.0	182.0
Black	229	0	0	229	76.4	23.6	0.0	183.8	0	0	229	89.2	89.2	1.3	174.5
Asian	21	0	0	21	47.6	52.4	0.0	198.7	0	0	21	57.1	57.1	28.6	200.2
Pacific Islander	6	0	0	6	50.0	50.0	0.0	204.7	0	0	6	0.0	0.0	202.2	202.2
Hispanic	100	0	0	100	73.0	27.0	0.0	185.9	0	0	100	87.1	87.1	12.9	176.0
Amer Indian/AK Native Other <sup>4</sup>	2	0	0	2	100.0	0.0	0.0	194.0	0	0	2	0.0	0.0	201.0	201.0
<b>Economic Status</b>															
Econ Disadvantaged	280	0	0	280	74.3	25.7	0.0	185.2	0	0	280	86.2	86.2	1.7	176.7
Non-Econ Disadvantaged	86	0	1	85	68.6	31.4	0.0	188.0	1	0	85	83.1	83.1	1.1	178.4

<sup>1</sup> Students appear in each applicable category, but they are included in Total Students only once.

<sup>2</sup> Excludes students coded as Former LEP.

<sup>3</sup> Excludes students who did not have Gender coded.

<sup>4</sup> Includes students who did not have an ethnicity coded and students who had more than one ethnicity coded.

**Note: The scores in this report are for illustrative purposes only.**

# **Appendix A**

## **Scoring Rubrics and 3rd Reader Score Calculation Charts**

# NEW JERSEY REGISTERED HOLISTIC SCORING RUBRIC

In scoring, consider the grid of written language	Inadequate Command	Limited Command	Partial Command	Adequate Command	Strong Command	Superior Command
Score	1	2	3	4	5	6
<b>Content and Organization</b>	<ul style="list-style-type: none"> <li>May lack opening and/or closing</li> <li>Minimal response to topic; uncertain focus</li> <li>No planning evident; disorganized</li> <li>Details random, inappropriate, or barely apparent</li> <li>No apparent control</li> <li>Severe/numerous errors</li> </ul>	<ul style="list-style-type: none"> <li>May lack opening and/or closing</li> <li>Attempts to focus</li> <li>May drift or shift focus</li> <li>Attempts organization</li> <li>Few, if any, transitions between ideas</li> <li>Details lack elaboration, i.e., highlight paper</li> <li>Numerous errors</li> <li>Excessive monotony/same structure</li> <li>Numerous errors</li> <li>Numerous serious errors</li> </ul>	<ul style="list-style-type: none"> <li>May lack opening and/or closing</li> <li>Usually has single focus</li> <li>Some lapses or flaws in organization</li> <li>May lack some transitions between ideas</li> <li>Repetitious details</li> <li>Several unelaborated details</li> <li>Errors/patterns of errors may be evident</li> <li>Little variety in syntax</li> <li>Some errors</li> <li>Patterns of errors evident</li> </ul>	<ul style="list-style-type: none"> <li>Generally has opening and/or closing</li> <li>Single focus</li> <li>Ideas loosely connected</li> <li>Transitions evident</li> <li>Uneven development of details</li> <li>Some errors that do not interfere with meaning</li> <li>Some variety</li> <li>Generally correct</li> <li>No consistent pattern of errors</li> <li>Some errors that do not interfere with meaning</li> </ul>	<ul style="list-style-type: none"> <li>Opening and closing</li> <li>Single focus</li> <li>Sense of unity and coherence</li> <li>Key ideas developed</li> <li>Logical progression of ideas</li> <li>Moderately fluent</li> <li>Attempts compositional risks</li> <li>Details appropriate and varied</li> <li>Few errors</li> <li>Variety in syntax appropriate and effective</li> <li>Few errors</li> <li>Few errors</li> </ul>	<ul style="list-style-type: none"> <li>Opening and closing</li> <li>Single, distinct focus</li> <li>Unified and coherent</li> <li>Well-developed</li> <li>Logical progression of ideas</li> <li>Fluent, cohesive</li> <li>Compositional risks successful</li> <li>Details effective, vivid, explicit, and/or pertinent</li> <li>Very few, if any, errors</li> <li>Precision and/or sophistication</li> <li>Very few, if any, errors</li> <li>Very few, if any, errors</li> </ul>
<b>Usage</b>						
<b>Sentence Construction</b>						
<b>Mechanics</b>						

	Content/Organization	Usage	Sentence Construction	Mechanics
<b>NON-SCORABLE RESPONSES</b>	<ul style="list-style-type: none"> <li>Communicates intended message to intended audience</li> <li>Relates to topic</li> <li>Opening and closing</li> <li>Focused</li> <li>Logical progression of ideas</li> <li>Transitions</li> <li>Appropriate details and information</li> </ul>	<ul style="list-style-type: none"> <li>Tense formation</li> <li>Subject-verb agreement</li> <li>Pronouns</li> <li>usage/agreement</li> <li>Word choice/meaning</li> <li>Proper Modifiers</li> </ul>	<ul style="list-style-type: none"> <li>Variety of type, structure, and length</li> <li>Correct construction</li> </ul>	<ul style="list-style-type: none"> <li>Spelling</li> <li>Capitalization</li> <li>Punctuation</li> </ul>
<b>NR = No Response</b>	Student wrote too little to allow a reliable judgment. of his/her writing.			
<b>OT = Off Topic/ Off Task</b>	Student did not write on the assigned topic/task, or the student attempted to copy the prompt.			
<b>NE = Not English</b>	Student wrote in a language other than English.			
<b>WF = Wrong Format</b>	Student refused to write on the topic, or the writing task folder was blank.			

Note: All unscorable responses, (NSRs), with the exception of NR, must be coded by the Scoring Director.

© New Jersey Department of Education

## OPEN-ENDED SCORING RUBRIC

### For Reading, Listening, and Viewing

**Sample Task:** The author takes a strong position on voting rights for young people. Use information from the text to support your response to the following.

- \*Requirements:**
- Explain the author’s position on voting.
  - Explain how adopting such a position would affect young people like you.

Points	Criteria
4	A 4-point response clearly demonstrates understanding of the task, completes all requirements, and provides an insightful explanation/opinion that links to or extends aspects of the text.
3	A 3-point response demonstrates an understanding of the task, completes all requirements, and provides some explanation/opinion using situations or ideas from the text as support.
2	A 2-point response may address all of the requirements, but demonstrates a partial understanding of the task, and uses text incorrectly or with limited success resulting in an inconsistent or flawed explanation.
1	A 1-point response demonstrates minimal understanding of the task, does not complete the requirements, and provides only a vague reference to or no use of the text.
0	A 0-point response is irrelevant or off-topic.

**\*Requirements for these items will vary according to the task.**

## REVISING/EDITING SCORING GUIDE

Score Point Scale	0	1	2	3	4
<b>Content/Organization</b> <ul style="list-style-type: none"> <li>● central focus</li> <li>● relevant supporting details</li> <li>● use of transitions and other devices to ensure cohesiveness</li> </ul>	no attention to opening or closing; no focus; no organization of ideas	minimal attention to opening or closing; some details but no elaboration; no transitions; unable to focus	limited attention to opening and/or closing; progression of ideas but flawed or uneven; may attempt to use transitions	general attention to opening and closing; sense of focus; some use of transitions, but uneven development may be overlooked	consistent attention to opening and closing; single, distinct focus; organization and elaboration of ideas; logical and cohesive use of transitions
<b>Sentence Construction</b> <ul style="list-style-type: none"> <li>● subordination/coordination</li> <li>● sentence fragments and run-on sentences</li> <li>● sentence combining</li> <li>● additional words to complete meaning</li> </ul>	incomplete/incorrect sentences	some sentence construction but marked by monotony and/or awkward syntax; no sense of rhetorical modes	some control of syntax; simple sentence structure, but little or no variety	control of syntax; eliminates excessive monotony; varied sentence structure	syntactic and rhetorical sophistication; subordination and coordination; avoids wordiness
<b>Usage</b> <ul style="list-style-type: none"> <li>● verbs (tense/agreement)</li> <li>● pronouns</li> <li>● (number/agreement)</li> <li>● parallel structure</li> <li>● correct modifiers</li> </ul>	numerous and/or serious errors ignored; inability to apply rules	some errors corrected but generally inconsistent application of rules	inconsistent in correcting errors; knowledge of rules but inability to utilize them effectively or consistently	errors corrected, but some may be overlooked; general knowledge and application of rules	knowledge and application of rules, leaving few, if any, errors
<b>Mechanics</b> <ul style="list-style-type: none"> <li>● spelling</li> <li>● punctuation</li> <li>● capitalization</li> </ul>	numerous and serious errors are ignored; inability to apply rules	inability to apply rules; errors, but inconsistently corrected; may create some errors where none existed	knowledge of rules, but some inconsistency in application; patterns of errors remain	errors, though some may be overlooked; knowledge and application of rules	errors, leaving few, if any; knowledge and application of rules
<b>Word choice</b>	no attention to word choice	limited word choice	relies on familiar vocabulary	varied vocabulary with some use of rich words	consistent use of rich words and images to develop topic

## Holistic Scoring Guide for Mathematics Open-Ended (OE) Items (Generic Rubric)

### 3-Point Response

The response shows complete understanding of the problem's essential mathematical concepts. The student executes procedures completely and gives relevant responses to all parts of the task. The response contains few minor errors, if any. The response contains a clear, effective explanation detailing how the problem was solved so that the reader does not need to infer how and why decisions were made.

### 2-Point Response

The response shows nearly complete understanding of the problem's essential mathematical concepts. The student executes nearly all procedures and gives relevant responses to most parts of the task. The response may have minor errors. The explanation detailing how the problem was solved may not be clear, causing the reader to make some inferences.

### 1-Point Response

The response shows limited understanding of the problem's essential mathematical concepts. The response and procedures may be incomplete and/or may contain major errors. An incomplete explanation of how the problem was solved may contribute to questions as to how and why decisions were made.

### 0-Point Response

The response shows insufficient understanding of the problem's essential mathematical concepts. The procedures, if any, contain major errors. There may be no explanation of the solution or the reader may not be able to understand the explanation. The reader may not be able to understand how and why decisions were made.

The above generic rubric is used as a guide to develop specific scoring guides or rubrics for each of the Open-Ended (OE) items which appear on the New Jersey eighth-grade (GEPA) and eleventh-grade (HSPA) proficiency assessments in Mathematics. The generic rubric helps insure that students are scored in the same way for the same demonstration of knowledge and skills regardless of the test question. More information on Open-Ended items and related scoring is also provided in the *Mathematics Instructional Guide*.

## HOLISTIC SCORING GUIDE (GENERIC RUBRIC) FOR SCIENCE OPEN-ENDED ITEMS AND PERFORMANCE ASSESSMENT TASKS

The zero-to-three-point generic scoring rubric below was created to help readers score open-ended responses consistently. In scoring, the reader should accept the use of appropriate diagrams, charts, formulas, and/or symbols which are part of a correct answer even when the question does not specifically request their use.

**3-Point Response:** Student response is reasonably complete, clear, and satisfactory.

**2-Point Response:** Student response has minor omissions and/or some incorrect or non-relevant information.

**1-Point Response:** Student response includes some correct information, but most information included in the response is either incorrect or not relevant.

**0-Point Response:** Student attempts the task but the response is incorrect, irrelevant, or inappropriate.

**Table 1**  
**Score Calculation Chart**  
 Used for Means ( $\bar{x}$ )  
 (Used for 3<sup>rd</sup> reading equal to or adjacent and all valid)

Absolute Difference ( $ 1^{\text{st}} - 2^{\text{nd}} $ )	Additional Conditions*	Additional Conditions*	Score Calculation*
0 <i>No 3<sup>rd</sup> Reading</i>	Both readings are valid	--	$(1^{\text{st}} + 2^{\text{nd}})/2$
1 <i>No 3<sup>rd</sup> Reading</i>	Both readings are valid	--	$(1^{\text{st}} + 2^{\text{nd}})/2$
2	$1^{\text{st}} < 3^{\text{rd}} < 2^{\text{nd}}$ or $2^{\text{nd}} < 3^{\text{rd}} < 1^{\text{st}}$	--	$(1^{\text{st}} + 2^{\text{nd}})/2$
2	$3^{\text{rd}} < ((1^{\text{st}} + 2^{\text{nd}})/2)$	$1^{\text{st}} < 2^{\text{nd}}$	$(1^{\text{st}} + 3^{\text{rd}})/2$
		$2^{\text{nd}} < 1^{\text{st}}$	$(2^{\text{nd}} + 3^{\text{rd}})/2$
	$3^{\text{rd}} > ((1^{\text{st}} + 2^{\text{nd}})/2)$	$1^{\text{st}} < 2^{\text{nd}}$	$(2^{\text{nd}} + 3^{\text{rd}})/2$
		$2^{\text{nd}} < 1^{\text{st}}$	$(1^{\text{st}} + 3^{\text{rd}})/2$
3	$3^{\text{rd}} = 1^{\text{st}}$ or $(3^{\text{rd}} \pm 1) = 1^{\text{st}}$	--	$(1^{\text{st}} + 3^{\text{rd}})/2$
	$3^{\text{rd}} = 2^{\text{nd}}$ or $(3^{\text{rd}} \pm 1) = 2^{\text{nd}}$	--	$(2^{\text{nd}} + 3^{\text{rd}})/2$
4 and 5	$3^{\text{rd}} = 1^{\text{st}}$ or $(3^{\text{rd}} \pm 1) = 1^{\text{st}}$	--	$(1^{\text{st}} + 3^{\text{rd}})/2$
	$3^{\text{rd}} = 2^{\text{nd}}$ or $(3^{\text{rd}} \pm 1) = 2^{\text{nd}}$	--	$(2^{\text{nd}} + 3^{\text{rd}})/2$

If both readings are invalid and equal, the score is 0

**Table 2**  
**Additional Score Calculations**

Used for Means ( $\chi$ )  
(Used for 3<sup>rd</sup> Reading NOT equal to or adjacent but all valid)

Condition	Score Calculation
1 <sup>st</sup> < 3 <sup>rd</sup> < 2 <sup>nd</sup> or 2 <sup>nd</sup> < 3 <sup>rd</sup> < 1 <sup>st</sup>	Use 3 <sup>rd</sup> reading
1 <sup>st</sup> < 2 <sup>nd</sup> < 3 <sup>rd</sup> or 3 <sup>rd</sup> < 2 <sup>nd</sup> < 1 <sup>st</sup>	$(2^{\text{nd}} + 3^{\text{rd}})/2$
2 <sup>nd</sup> < 1 <sup>st</sup> < 3 <sup>rd</sup> or 3 <sup>rd</sup> < 1 <sup>st</sup> < 2 <sup>nd</sup>	$(1^{\text{st}} + 3^{\text{rd}})/2$

**Table 2A**  
Used for Means ( $\chi$ )  
(Used for 1<sup>st</sup> or 2<sup>nd</sup> reading invalid and 3<sup>rd</sup> Reading valid)

Condition	Additional Condition	Score Calculation
1 <sup>st</sup> Reading Invalid 2 <sup>nd</sup> Reading Valid	Absolute difference between 2 <sup>nd</sup> Reading and 3 <sup>rd</sup> reading is 0 or 1	$(2^{\text{nd}} + 3^{\text{rd}}) / 2$
	Absolute difference between 2 <sup>nd</sup> Reading and 3 <sup>rd</sup> reading is greater than 1	Use 3 <sup>rd</sup> Reading
1 <sup>st</sup> Reading Valid 2 <sup>nd</sup> Reading Invalid	Absolute difference between 1 <sup>st</sup> Reading and 3 <sup>rd</sup> reading is 0 or 1	$(1^{\text{st}} + 3^{\text{rd}}) / 2$
	Absolute difference between 1 <sup>st</sup> Reading and 3 <sup>rd</sup> reading is greater than 1	Use 3 <sup>rd</sup> Reading
Both 1 <sup>st</sup> and 2 <sup>nd</sup> Readings are invalid		Use 3 <sup>rd</sup> Reading

If the 3<sup>rd</sup> Reading is invalid, use the 3<sup>rd</sup> reading score

**Table 3**  
**Score Calculation Chart**

Used for Sum ( $\Sigma$ )  
(Used for 3<sup>rd</sup> Reading equal to or adjacent and all valid)

Absolute Difference ( $ 1^{\text{st}} - 2^{\text{nd}} $ )	Additional Conditions*	Additional Conditions*	Score Calculation*
0 <i>No 3<sup>rd</sup> Reading</i>	--	--	$(1^{\text{st}} + 2^{\text{nd}})$
1 <i>No 3<sup>rd</sup> Reading</i>	--	--	$(1^{\text{st}} + 2^{\text{nd}})$
2 – 5	Equal to or Adjacent	--	$((1^{\text{st}} + 2^{\text{nd}} + 3^{\text{rd}}) * 2) / 3$

**Table 4****Additional Score Calculations**Used for Sum ( $\Sigma$ )(Used for 3<sup>rd</sup> Reading NOT equal to or adjacent but all valid)

Condition	Score Calculation
NOT Equal to or Adjacent	$((1^{\text{st}} + 2^{\text{nd}} + 3^{\text{rd}}) * 2) / 3$

If the 3<sup>rd</sup> Reading is invalid, use the 3<sup>rd</sup> reading score**Table 4A**Used for Sum ( $\Sigma$ )(Used for 1<sup>st</sup> or 2<sup>nd</sup> reading invalid and 3<sup>rd</sup> Reading valid)

Condition	Additional Condition	Score Calculation
1 <sup>st</sup> Reading Invalid 2 <sup>nd</sup> Reading Valid	Absolute difference between 2 <sup>nd</sup> Reading and 3 <sup>rd</sup> reading is 0 or 1	$(2^{\text{nd}} + 3^{\text{rd}})$
	Absolute difference between 2 <sup>nd</sup> Reading and 3 <sup>rd</sup> reading is greater than 1	$(3^{\text{rd}} * 2)$
1 <sup>st</sup> Reading Valid 2 <sup>nd</sup> Reading Invalid	Absolute difference between 1 <sup>st</sup> Reading and 3 <sup>rd</sup> reading is 0 or 1	$(1^{\text{st}} + 3^{\text{rd}})$
	Absolute difference between 1 <sup>st</sup> Reading and 3 <sup>rd</sup> reading is greater than 1	$(3^{\text{rd}} * 2)$
Both 1 <sup>st</sup> and 2 <sup>nd</sup> Readings are invalid		Use 3 <sup>rd</sup> Reading

If the 3<sup>rd</sup> Reading is invalid, use the 3<sup>rd</sup> reading score

**Table 5**  
**When to Use the Mean vs. Sum Scoring Rules**

Subject	Valid scores	Grade 8
Reading OE	0-4 *	Mean
Writing – Picture	Grade 8 1-6 **	Mean
Writing – Persuasive	1-6 **	Sum
Revise / Edit	0-4 *	Sum
Math OE	0-3 *	Mean
Science OE	0-3 *	Mean

Designation Codes:

- \* = 7 = NR, for No Response  
(blank, fragmented, refusing or unable to write on topic, copy of item)  
8 = OT, for Off Topic  
9 = NE, for Not English
- \*\* = 0 = NR, for No Response  
7 = WF, for Wrong Format  
8 = OT, for Off Topic  
9 = NE, for Not English
- \*\*\* = 7 = NR, No Response

There are three types of situations that will require a third reading:

1. First and second reading are valid scores and not equal or adjacent.
2. One reading is a valid score and the other reading is not a valid score
3. Both readings are not a valid score and are not equal.

# ***Appendix B***

## ***Cycle II Test Results***

**Executive Summary**

The 2004 New Jersey Grade Eight Proficiency Assessment (GEPA) consisted of three sections: Language Arts Literacy, Mathematics, and Science. The GEPA is to be used as a primary indicator for identifying those students who may need instructional intervention in the three content areas. It is designed to give an indication of the progress students are making in mastering the skills they will need to pass the High School Proficiency Assessment (HSPA).

The total GEPA Language Arts Literacy, Mathematics, and Science scores are reported as scale scores with a range of 100 to 300. Please note that 100 and 300 are a theoretical floor and ceiling, which may not actually be observed. The score ranges are as follows:

---

<i>Advanced Proficient</i>	<i>250-300</i>
<i>Proficient</i>	<i>200-249</i>
<i>Partially Proficient</i>	<i>100-199</i>

---

Students who scored Partially Proficient are considered to be below the state minimum level of proficiency. These students may need additional instructional support such as individual or programmatic intervention. It is important that districts consider multiple measures with all students before making the final decisions about students’ instructional placement.

The Statewide Performance by Demographic Groups Report shows enrollment and performance data for various demographic groups in New Jersey. For each demographic group, the number of students participating, the percent of students in each proficiency level, and the mean scale score are reported for each content area.

Students are counted in the “Total Students” category only once, but are counted in other categories that apply. The report groups students by instructional groups, gender, migrant status, economic status, and ethnicity. Students may be counted in as many ethnicities as they grid on their answer documents. The demographic information originates from the data collected on the students’ answer folders. These data are reviewed by the school districts prior to reporting, allowing them to correct any errors.

Following is an explanation of how to interpret the categories of students presented in the report. “Total students” refer to all students tested. “General Education” excludes students coded as special education OR limited English proficient (LEP) on their answer folders. “Special Education” (SE) includes students coded as SE. “Limited English Proficient” includes students coded as LEP. The performance data only include students who received valid scale scores. Students whose answer folders were voided were excluded. Because each content area is independent, students may receive a scale score in one content area, but not in others.

### Highlights from the Performance by Demographic Groups Report – State

The GEPA was administered between March 8 and March 11, 2004. The test was administered to 110,270 students. Of these, 108,427 had valid scale scores in Language Arts Literacy, 108,965 had valid scale scores in Mathematics and 108,841 students had valid scale scores in science.

For **Total Students**, 28.3% scored Partially Proficient, 66.3% Proficient; and 5.5% Advanced Proficient in Language Arts Literacy. In Mathematics, 38.3% were Partially Proficient, 41.7% Proficient and 20% Advanced Proficient. In Science, 25.3% were Partially Proficient, 53.6% Proficient and 21.1% Advanced Proficient. The mean scale score was 211.9 in Language Arts Literacy, 212.6 in Mathematics and 222.2 in Science.

For **Special Education (SE) students**, 72.3% were Partially Proficient, 27.4% were Proficient and 0.2% Advanced Proficient in the Language Arts Literacy. In Mathematics, 79.2% scored Proficient, 18.4% were Proficient and 2.4% were Advanced Proficient. In Science, 55.9% were Partially Proficient, 39.3% were Proficient and 4.8% Advanced Proficient. The mean scale score was 181.4 in Language Arts Literacy, 180.1 in Mathematics and 199 in Science.

For **Limited English Proficient Students**, 82.5% were Partially Proficient, 17.3% were Proficient and 0.2% were Advanced Proficient in the Language Arts Literacy. In Mathematics, 76.5% scored Partially Proficient, 19.% were Proficient and 4.6% were Advanced Proficient. In Science, 73.9% were Partially Proficient, 24.7% were Proficient and 1.4% were Advanced Proficient. The mean scale score was 171 in Language Arts Literacy, 182.4 in Mathematics and 188.2 in Science.

As far as **Gender** is concerned, in the Language Arts Literacy, 22% of females were Partially Proficient, 70.1% were Proficient, and 8% were Advanced Proficient compared to 34.2%, 62.7%, and 3.1% of males scoring Partially Proficient, Proficient, and Advanced Proficient respectively. In Mathematics, 38.8% of females were Partially Proficient, 42.8% were Proficient, and 18.4% were Advanced Proficient compared to 37.8%, 40.6%, and 21.6% of males scoring Partially Proficient, Proficient and Advanced Proficient respectively. In Science, 27.7% of females were Partially Proficient, 55.3% were Proficient, and 17% were Advanced Proficient compared to 22.9%, 52%, and 25% of males scoring Partially Proficient, Proficient, and Advanced Proficient respectively. The mean scale score was 217.6 in Language Arts Literacy, 211.7 in Mathematics and 219.1 in Science for females and 206.4, 213.4 and 225.1 for males.

Of different **Ethnicities**, the percentages of Proficient and Advanced Proficient ranged from 85.8% Asian students to 46.5% for Black students in Language Arts Literacy. In Mathematics, the percentages of Proficient and Advanced Proficient students ranged from 82.9 for Asian students to 30.3% for Black students. In Science, Proficient and Advanced Proficient percentages ranged from 86.9% for White students to 47.9% for Black students.

For **Economically Disadvantaged** students, 53.4% scored Proficient, 45.7% Proficient and 0.9% Advanced Proficient in Language Arts Literacy compared to 19.3%, 73.6%, and 7.1% of Non-Economically Disadvantaged students scoring Partially Proficient, Proficient and Advanced Proficient respectively. In Mathematics, 63.9% of Economically Disadvantaged students scored Partially Proficient, 30.2% Proficient and 5.9% Advanced Proficient compared to 29.1%, 45.8%, and 25.1% of Non-Economically Disadvantaged students in similar proficiency levels. In Science, 49.9% of Economically Disadvantaged students scored Partially Proficient, 44.7% scored Proficient, and 5.4% scored Advanced Proficient compared to 16.4%, 56.8%, and 26.8% of Non-Economically Disadvantaged students in the similar proficiency levels respectively. The mean scale score was 193.9 in Language Arts Literacy, 191.7 in Mathematics and 202.4 in Science for Economically Disadvantaged students compared to 218.2, 220 and 229.2 for Non-Economically Disadvantaged students.

### ***Reporting Rules for State Summary Data File***

The accompanying state summary data file contains the same type of information shown on the Cycle II Statewide Performance by Demographic Groups Report for by schools and districts. In order to safeguard confidentiality, certain information is suppressed in the file according to the following reporting rules:

- Data are not reported where the number of students with valid scale scores for a particular group is less than 11.
- Data are not reported where demographic groups are mutually exclusive (e.g., gender) and there are one or two students with a valid scale score in one of the groups (e.g., male).
- Data are not reported when it is otherwise possible to identify individual students' performance.



**New Jersey Statewide Testing System  
Grade Eight Proficiency Assessment  
Performance by Demographic Groups  
State**

Test Date: March 2004  
CYCLE II

	Language Arts Literacy										Mathematics										Science									
	Number Enrolled	Number Not Present	Number of Voids	Number of Valid Scale Scores	% Partially Proficient	% Proficient	Advanced Proficient	Scale Score Mean	Number Not Present	Number of Voids	Number of Valid Scale Scores	% Partially Proficient	% Proficient	Advanced Proficient	Scale Score Mean	Number Not Present	Number of Voids	Number of Valid Scale Scores	% Partially Proficient	% Proficient	Advanced Proficient	Scale Score Mean								
<b>Total Students<sup>1</sup></b>	110,270	564	1,299	108,427	28.3	66.3	5.5	211.9	687	618	108,965	38.3	41.7	20.0	212.6	886	543	106,841	25.3	53.6	21.1	222.2								
General Education	88,480	277	424	87,779	17.5	75.8	6.7	219.4	553	82	88,045	28.7	47.2	24.1	220.2	446	83	87,951	17.3	57.6	25.1	228.1								
Special Education	18,010	230	837	16,943	72.3	27.4	0.2	181.4	296	530	17,184	79.2	18.4	2.4	180.1	398	448	17,164	55.9	39.3	4.8	199.0								
Limited English Proficient	3,982	39	50	3,893	82.5	17.3	0.2	171.0	40	14	3,928	76.5	19.0	4.6	182.4	47	19	3,916	73.9	24.7	1.4	186.2								
<b>Gender<sup>2</sup></b>																														
Female	55,604	218	469	52,917	22.0	70.1	8.0	217.6	289	222	53,093	36.8	42.8	18.4	211.7	362	200	53,042	27.7	55.3	17.0	219.1								
Male	56,542	315	821	55,406	34.2	62.7	3.1	206.4	386	390	55,766	37.8	40.6	21.6	213.4	511	337	55,694	22.9	52.0	25.0	225.1								
<b>Migrant Status</b>																														
Migrant	74	1	3	70	60.0	37.1	2.9	191.7	1	2	71	73.2	21.1	5.6	186.9	3	0	71	53.5	43.7	2.8	199.1								
Non-Migrant	110,196	543	1,296	108,357	28.2	66.3	5.5	211.9	686	616	108,894	38.3	41.7	20.0	212.6	883	543	108,770	25.2	53.6	21.1	222.2								
<b>Ethnicity<sup>3</sup></b>																														
White	64,697	165	539	63,993	16.8	76.1	7.1	219.8	227	307	64,163	25.7	48.2	26.0	222.4	303	267	64,127	13.2	58.2	28.6	231.7								
Black	19,758	215	419	19,124	53.5	45.6	0.8	194.5	271	162	19,325	69.7	26.6	3.6	187.6	336	146	19,276	52.1	43.5	4.3	201.0								
Asian	6,749	8	42	6,699	14.2	71.3	14.5	224.8	8	30	6,711	17.1	40.3	42.6	234.5	13	28	6,708	13.1	50.8	36.0	235.2								
Pacific Islander	472	1	3	468	17.5	74.8	7.7	220.4	2	0	470	21.7	45.1	33.2	226.7	2	0	470	14.7	57.7	27.7	230.6								
Hispanic	17,935	139	265	17,531	47.8	50.9	1.3	196.8	163	110	17,662	57.6	35.0	7.4	195.9	202	93	17,640	44.4	49.3	6.3	205.7								
Amer Indian/AK Native	172	0	4	168	31.0	66.1	3.0	207.9	0	2	170	41.2	41.2	17.6	207.7	0	2	170	28.8	53.5	17.6	217.9								
<b>Economic Status</b>																														
Econ Disadvantaged	29,300	281	555	28,464	53.4	45.7	0.9	193.9	347	209	28,744	63.9	30.2	5.9	191.7	427	202	28,671	49.9	44.7	5.4	202.4								
Non-Econ Disadvantaged	80,970	263	744	79,963	19.3	73.6	7.1	218.2	340	409	80,221	29.1	45.8	25.1	220.0	459	341	80,170	16.4	56.8	26.8	229.2								

<sup>1</sup> Students appear in each applicable category, but they are included in Total Students only once.  
<sup>2</sup> Excludes students who did not have Gender coded.  
<sup>3</sup> Excludes students who did not have an ethnicity coded. Students with multiple ethnicities are reported for all ethnic categories that apply.



## New Jersey Statewide Testing System Grade Eight Proficiency Assessment Performance by Demographic Groups DFG A

Test Date: March 2004  
CYCLE II

	Language Arts Literacy						Mathematics						Science										
	Number Enrolled	Number Not Present <sup>1</sup>	Number of Valid Scores	% Partially Proficient	% Proficient	Advanced Proficient	Scale Score Mean	Number Not Present <sup>1</sup>	Number of Valid Scores	% Partially Proficient	% Proficient	Advanced Proficient	Scale Score Mean	Number Not Present <sup>1</sup>	Number of Valid Scores	% Partially Proficient	% Proficient	Advanced Proficient	Scale Score Mean				
<b>Total Students<sup>1</sup></b>	19,398	219	431	18,748	57.6	41.5	0.9	191.2	286	154	18,958	66.8	27.5	5.7	189.5	338	137	18,923	56.8	39.5	3.7	198.3	
General Education	13,606	114	153	13,339	44.8	53.9	1.3	201.4	150	24	13,432	58.1	34.3	7.6	196.9	169	22	13,415	46.6	48.3	5.0	204.2	
Special Education	3,857	82	256	3,519	91.5	8.5	0.0	164.5	110	129	3,618	92.2	7.4	0.4	168.1	142	109	3,606	83.3	16.3	0.4	182.8	
Limited English Proficient	2,072	24	32	2,016	85.3	14.7	0.0	167.9	27	9	2,036	80.4	17.4	2.2	178.2	31	12	2,029	78.2	21.0	0.8	185.6	
<b>Gender<sup>2</sup></b>																							
Female	9,575	96	157	9,322	50.0	48.5	1.5	197.2	118	59	9,398	66.1	28.5	5.4	190.2	139	49	9,387	59.2	37.9	2.9	196.7	
Male	9,797	120	272	9,405	65.1	34.6	0.4	185.2	164	95	9,538	67.5	26.5	6.0	188.9	195	87	9,515	54.3	41.1	4.6	199.9	
<b>Migrant Status</b>																							
Migrant	31	1	29	29	79.3	20.7	0.0	178.5	1	1	29	86.2	10.3	3.4	176.7	2	0	29	58.6	37.9	3.4	192.3	
Non-Migrant	19,367	218	430	18,719	57.6	41.5	0.9	191.2	285	153	18,929	66.8	27.5	5.7	189.6	336	137	18,894	56.8	39.5	3.7	198.3	
<b>Ethnicity<sup>3</sup></b>																							
White	1,571	13	24	1,534	36.3	60.9	2.8	205.2	16	12	1,543	42.2	43.6	14.2	207.9	19	16	1,536	31.4	56.3	12.4	214.6	
Black	8,731	119	228	8,384	64.2	35.4	0.4	187.8	157	70	8,504	76.7	20.8	2.5	182.4	190	59	8,482	65.3	33.0	1.7	193.2	
Asian	510	2	5	503	36.0	60.8	3.2	205.1	3	5	502	40.8	38.8	20.3	211.5	3	5	502	35.7	50.0	14.3	213.8	
Pacific Islander	25	0	0	25	32.0	68.0	0.0	206.0	0	0	25	40.0	40.0	20.0	212.0	0	0	25	32.0	56.0	12.0	215.6	
Hispanic	8,484	82	164	8,238	55.9	43.2	0.9	191.4	105	66	8,313	62.7	30.8	6.5	192.2	119	53	8,312	53.8	42.6	3.6	199.6	
Amer Indian/AK Native	49	0	0	49	49.0	51.0	0.0	199.4	0	0	49	53.1	30.6	16.3	200.2	0	0	49	49.0	40.8	10.2	206.6	
<b>Economic Status</b>																							
Econ Disadvantaged	14,427	151	316	13,960	59.4	39.9	0.7	190.1	199	113	14,115	68.1	26.7	5.2	188.6	238	104	14,085	58.6	3.3	3.2	197.0	
Non-Econ Disadvantaged	4,971	68	115	4,788	52.5	46.1	1.4	194.3	87	41	4,843	63.2	29.5	7.2	192.2	100	33	4,838	51.5	43.1	5.4	202.1	

<sup>1</sup> Students appear in each applicable category, but they are included in Total Students only once.

<sup>2</sup> Excludes students who did not have Gender coded.

<sup>3</sup> Excludes students who did not have an ethnicity coded. Students with multiple ethnicities are reported for all ethnic categories that apply.



**New Jersey Statewide Testing System  
Grade Eight Proficiency Assessment  
Performance by Demographic Groups  
DFG B**

Test Date: March 2004  
CYCLE II

	Language Arts Literacy										Mathematics										Science									
	Number Enrolled	Number Not Present	Number of Voids	Number of Valid Scale Scores	% Proficient	% Partially Proficient	% Advanced Proficient	Scale Score Mean	Number Not Present	Number of Voids	Number of Valid Scale Scores	% Proficient	% Partially Proficient	% Advanced Proficient	Scale Score Mean	Number Not Present	Number of Voids	Number of Valid Scale Scores	% Proficient	% Partially Proficient	% Advanced Proficient	Scale Score Mean								
<b>Total Students<sup>1</sup></b>	11,402	102	157	11,143	39.0	58.9	2.0	203.1	117	79	11,206	49.8	39.6	10.7	202.5	71	11,194	33.2	55.5	11.3	213.8									
General Education	8,609	38	53	8,518	25.1	72.3	2.7	212.6	45	9	8,555	38.9	47.5	13.6	210.7	56	8,544	22.9	62.8	14.3	220.4									
Special Education	2,260	61	102	2,097	83.2	16.8	0.0	174.2	69	69	2,122	85.9	13.1	1.0	175.0	77	2,122	63.4	34.6	2.0	193.9									
Limited English Proficient	550	4	3	543	89.1	10.9	0.0	165.6	4	1	545	81.7	16.7	1.7	179.4	5	544	78.3	21.1	0.6	186.3									
<b>Gender<sup>2</sup></b>																														
Female	5,501	31	50	5,420	30.9	66.0	3.0	209.2	41	25	5,435	50.8	40.0	9.2	201.8	48	5,425	36.7	55.3	8.0	210.5									
Male	5,891	70	106	5,715	46.8	52.2	1.1	197.4	75	53	5,763	48.9	39.1	12.0	203.2	88	5,761	29.9	55.7	14.4	216.9									
<b>Migrant Status</b>																														
Migrant	3	0	0	3	66.7	33.3	0.0	156.3	0	0	3	66.7	33.3	0.0	181.0	0	3	66.7	33.3	0.0	196.0									
Non-Migrant	11,399	102	157	11,140	39.0	58.9	2.0	203.1	117	79	11,203	49.8	39.6	10.7	202.5	137	11,191	33.2	55.5	11.3	213.8									
<b>Ethnicity<sup>3</sup></b>																														
White	5,905	32	61	5,812	30.7	66.6	2.8	208.6	40	32	5,833	40.4	45.2	14.4	209.6	42	5,830	23.4	60.5	16.1	220.9									
Black	2,094	38	49	2,007	53.6	45.8	0.6	194.6	47	26	2,021	71.0	26.1	2.9	187.0	51	2,022	50.7	45.9	3.4	201.4									
Asian	276	0	2	274	28.1	67.5	4.4	210.5	0	1	275	30.5	45.8	23.6	220.1	0	275	25.5	56.0	18.5	220.3									
Pacific Islander	46	0	0	46	19.6	73.9	6.5	219.0	0	0	46	26.1	47.8	26.1	219.0	0	46	15.2	60.9	23.9	228.1									
Hispanic	3,051	32	44	2,975	46.9	51.8	1.3	197.4	31	19	3,001	55.6	37.2	7.1	197.2	44	2,992	41.2	52.2	6.6	207.5									
Amer Indian/AK Native	18	0	1	17	52.9	47.1	0.0	190.0	0	1	17	47.1	35.3	17.6	203.2	0	17	35.3	52.9	11.8	211.7									
<b>Economic Status</b>																														
Econ Disadvantaged	4,564	54	82	4,428	51.5	47.8	0.7	194.7	63	39	4,462	61.6	32.7	5.7	193.0	68	4,461	45.4	48.4	6.2	205.3									
Non-Econ Disadvantaged	6,838	48	75	6,715	30.8	66.3	2.9	208.7	54	40	6,744	41.9	44.1	13.9	208.7	69	6,733	25.0	60.3	14.7	219.4									

<sup>1</sup> Students appear in each applicable category, but they are included in Total Students only once.

<sup>2</sup> Excludes students who did not have Gender coded.

<sup>3</sup> Excludes students who did not have an ethnicity coded. Students with multiple ethnicities are reported for all ethnic categories that apply.



## New Jersey Statewide Testing System Grade Eight Proficiency Assessment Performance by Demographic Groups DFG CD

Test Date: March 2004  
CYCLE II

	Language Arts Literacy										Mathematics										Science									
	Number Not Present	Number of Votes	Number of Valid Scale Scores	% Proficient	% Partially Proficient	% Advanced Proficient	Scale Score Mean	Number Not Present	Number of Voids	Number of Valid Scale Scores	% Proficient	% Partially Proficient	% Advanced Proficient	Scale Score Mean	Number Not Present	Number of Voids	Number of Valid Scale Scores	% Proficient	% Partially Proficient	% Advanced Proficient	Scale Score Mean	Number Not Present	Number of Voids	Number of Valid Scale Scores	% Proficient	% Partially Proficient	% Advanced Proficient	Scale Score Mean		
<b>Total Students<sup>1</sup></b>	68	136	10,322	31.9	65.1	3.0	208.4	81	62	10,385	45.4	80.2	16.3	180.5	3	1	263	80.2	16.3	3.4	180.5	3	1	263	77.6	21.7	0.8	186.5		
General Education	39	37	8,258	20.3	76.0	3.7	216.0	44	9	8,261	35.5	84.4	1.2	176.2	62	11	8,241	17.8	63.6	15.4	212.7	62	11	8,241	17.8	63.6	15.4	212.7		
Special Education	26	97	1,835	77.2	22.7	0.1	179.3	34	52	1,872	84.4	16.3	3.4	180.5	54	45	1,859	57.1	39.9	1.2	176.2	54	45	1,859	57.1	39.9	1.2	176.2		
Limited English Proficient	267	3	260	83.8	16.2	0.0	171.2	3	1	263	80.2	16.3	3.4	180.5	3	1	263	80.2	16.3	3.4	180.5	3	1	263	77.6	21.7	0.8	186.5		
<b>Gender<sup>2</sup></b>																														
Female	33	47	4,969	22.6	72.7	4.7	215.2	40	17	4,972	45.6	80.2	16.3	180.5	52	18	4,959	29.4	58.6	11.6	205.1	52	18	4,959	29.4	58.6	11.6	205.1		
Male	34	90	5,360	40.5	58.1	1.4	202.1	40	44	5,400	45.2	80.2	16.3	180.5	66	38	5,380	23.5	58.3	13.4	205.7	66	38	5,380	23.5	58.3	13.4	205.7		
<b>Migrant Status</b>																														
Migrant	4	1	3	66.7	33.3	0.0	192.7	0	1	3	100.0	0.0	0.0	181.7	1	0	3	66.7	33.3	0.0	181.7	1	0	3	66.7	33.3	0.0	181.7		
Non-Migrant	68	137	10,319	31.9	65.1	3.0	208.4	81	61	10,382	45.3	80.2	16.3	180.5	118	57	10,349	26.3	58.4	12.6	205.4	118	57	10,349	26.3	58.4	12.6	205.4		
<b>Ethnicity<sup>2</sup></b>																														
White	33	79	6,311	23.3	72.8	3.9	214.2	41	37	6,345	35.0	80.2	16.3	180.5	57	34	6,332	17.3	62.3	16.7	213.2	57	34	6,332	17.3	62.3	16.7	213.2		
Black	23	43	2,162	49.0	49.9	1.1	197.3	28	19	2,161	69.1	28.0	2.9	187.3	43	18	2,147	44.8	50.2	2.9	187.3	43	18	2,147	44.8	50.2	2.9	187.3		
Asian	0	1	415	28.4	67.0	4.6	211.7	0	1	415	27.0	48.4	24.6	220.9	0	1	415	19.8	60.0	24.6	220.9	0	1	415	19.8	60.0	24.6	220.9		
Pacific Islander	0	0	15	33.3	66.7	0.0	213.3	0	0	15	40.0	33.3	26.7	218.3	0	0	15	13.3	60.0	26.7	218.3	0	0	15	13.3	60.0	26.7	218.3		
Hispanic	10	12	1,385	46.4	52.6	1.0	197.8	10	5	1,392	61.3	33.8	5.0	193.1	15	4	1,388	41.1	52.4	5.0	193.1	15	4	1,388	41.1	52.4	5.0	193.1		
Amer Indian/AK Native	12	0	12	8.3	83.3	8.3	221.4	0	0	12	58.3	33.3	8.3	196.1	0	0	12	16.7	75.0	8.3	196.1	0	0	12	16.7	75.0	8.3	196.1		
<b>Economic Status</b>																														
Econ Disadvantaged	32	49	2,964	49.1	49.9	1.0	196.9	34	16	2,995	63.5	31.2	5.3	191.8	55	18	2,972	42.2	51.5	5.3	191.8	55	18	2,972	42.2	51.5	5.3	191.8		
Non-Econ Disadvantaged	36	89	7,358	25.0	71.2	3.8	213.0	47	46	7,390	38.0	46.5	15.5	210.9	64	39	7,380	19.9	61.1	15.5	210.9	64	39	7,380	19.9	61.1	15.5	210.9		

<sup>1</sup> Students appear in each applicable category, but they are included in Total Students only once.

<sup>2</sup> Excludes students who did not have Gender coded.

<sup>3</sup> Excludes students who did not have an ethnicity coded. Students with multiple ethnicities are reported for all ethnic categories that apply.



**New Jersey Statewide Testing System  
Grade Eight Proficiency Assessment  
Performance by Demographic Groups  
DFG DE**

Test Date: March 2004  
CYCLE II

	Number Enrolled	Language Arts Literacy					Mathematics					Science										
		Number Not Present	Number of Voids	Number of Valid Scale Scores	% Partially Proficient	% Advanced Proficient	Scale Score Mean	Number Not Present	Number of Voids	Number of Valid Scale Scores	% Partially Proficient	% Advanced Proficient	Scale Score Mean	Number Not Present	Number of Voids	Number of Valid Scale Scores	% Partially Proficient	% Advanced Proficient	Scale Score Mean			
<b>Total Students<sup>1</sup></b>	17,259	51	174	17,014	25.2	71.0	3.8	213.2	71	83	17,085	36.4	46.3	17.3	212.7	94	68	17,077	20.8	59.1	20.1	224.1
General Education	14,292	28	73	14,191	15.4	80.0	4.6	219.4	37	12	14,243	27.8	51.8	20.4	219.2	45	14	14,233	14.0	62.7	23.4	229.0
Special Education	2,623	21	98	2,504	73.5	26.4	0.1	182.7	33	70	2,520	80.6	17.9	1.5	179.5	48	53	2,522	53.8	42.2	3.9	200.3
Limited English Proficient	358	2	3	353	79.3	20.1	0.6	176.0	1	1	356	72.0	24.7	3.3	184.9	1	1	356	67.0	32.1	0.9	191.0
<b>Gender<sup>2</sup></b>																						
Female	8,378	18	64	8,296	17.8	76.4	5.8	219.1	33	26	8,319	36.9	47.0	16.1	212.1	35	23	8,320	23.0	60.8	16.2	221.1
Male	8,852	32	107	8,693	32.1	66.0	1.9	207.5	37	55	8,740	35.8	45.6	18.5	213.4	58	43	8,731	18.7	57.5	23.8	227.0
<b>Migrant Status</b>																						
Migrant	34	0	1	33	45.5	48.5	6.1	205.4	0	0	34	58.8	32.4	8.8	197.1	0	0	34	47.1	50.0	2.9	205.4
Non-Migrant	17,205	51	173	16,981	25.1	71.1	3.8	213.2	71	83	17,051	36.4	46.3	17.3	212.8	94	68	17,043	20.8	59.1	20.1	224.1
<b>Ethnicity<sup>3</sup></b>																						
White	11,868	29	96	11,743	20.2	75.4	4.4	216.6	48	47	11,773	30.0	49.6	20.4	217.6	60	34	11,774	15.5	60.4	24.1	228.6
Black	2,199	5	41	2,153	41.9	56.9	1.1	201.6	9	21	2,169	63.1	32.8	4.1	192.9	18	19	2,162	39.0	53.8	7.2	208.9
Asian	928	2	8	918	17.5	74.6	7.8	220.1	0	5	923	20.8	49.2	30.0	226.5	2	5	921	16.9	56.9	26.2	228.8
Pacific Islander	189	0	2	187	20.9	74.9	4.3	217.6	0	0	189	23.8	48.7	27.5	223.0	0	0	189	15.9	65.1	19.0	226.5
Hispanic	1,960	10	20	1,930	39.4	59.0	1.6	202.7	10	8	1,942	52.2	40.4	7.4	199.4	10	9	1,941	32.9	58.8	8.3	212.6
Amer Indian/AK Native	35	0	1	34	26.5	73.5	0.0	208.8	0	0	35	34.3	57.1	8.6	207.9	0	0	35	25.7	60.0	14.3	216.1
<b>Economic Status</b>																						
Econ Disadvantaged	2,944	13	42	2,889	44.2	54.4	1.4	200.6	19	11	2,914	55.6	36.3	8.0	198.0	27	12	2,905	35.8	54.9	9.2	211.3
Non-Econ Disadvantaged	14,295	38	132	14,125	21.3	74.4	4.3	215.7	52	72	14,171	32.5	48.3	19.2	215.8	67	56	14,172	17.8	59.9	22.3	226.7

<sup>1</sup> Students appear in each applicable category, but they are included in Total Students only once.

<sup>2</sup> Excludes students who did not have Gender coded.

<sup>3</sup> Excludes students who did not have an ethnicity coded. Students with multiple ethnicities are reported for all ethnic categories that apply.



## New Jersey Statewide Testing System Grade Eight Proficiency Assessment Performance by Demographic Groups DFG FG

Test Date: March 2004  
CYCLE II

	Language Arts Literacy						Mathematics						Science										
	Number Enrolled	Number Not Present	Number of Valid Scale Scores	% Proficient	% Partially Proficient	Advanced Proficient	Scale Score Mean	% Proficient	% Partially Proficient	Advanced Proficient	Scale Score Mean	Number of Valid Scale Scores	% Proficient	% Partially Proficient	Advanced Proficient	Scale Score Mean							
<b>Total Students<sup>1</sup></b>	13,724	21	126	13,577	20.5	74.2	5.3	216.3	34	67	13,623	31.0	46.9	22.1	217.8	69	61	13,594	16.1	59.3	24.6	228.2	
General Education	11,440	11	34	11,395	11.1	82.6	6.3	222.2	20	5	11,415	22.1	52.0	25.9	224.5	40	6	11,394	9.9	61.6	28.6	233.1	
Special Education	2,075	8	90	1,977	68.7	31.2	0.1	186.9	10	62	2,003	77.4	20.6	1.9	182.7	26	55	1,994	45.8	49.7	4.5	204.1	
Limited English Proficient	219	2	3	214	80.4	19.6	0.0	172.3	4	0	215	72.6	19.1	8.4	187.4	3	1	215	71.2	27.0	1.9	188.6	
<b>Gender<sup>2</sup></b>																							
Female	6,741	8	48	6,685	14.8	77.5	7.7	221.5	15	27	6,699	31.8	48.2	20.0	216.5	35	26	6,680	18.6	61.9	19.4	224.5	
Male	6,978	13	78	6,887	26.1	71.0	3.0	211.2	19	40	6,919	30.2	45.7	24.1	219.1	34	35	6,909	13.6	56.8	29.6	231.9	
<b>Migrant Status</b>																							
Migrant																							
Non-Migrant	13,724	21	126	13,577	20.5	74.2	5.3	216.3	34	67	13,623	31.0	46.9	22.1	217.8	69	61	13,594	16.1	59.3	24.6	228.2	
<b>Ethnicity<sup>3</sup></b>																							
White	10,176	16	86	10,074	17.4	77.3	5.3	218.2	26	46	10,104	27.4	49.3	23.3	220.4	51	41	10,084	12.6	60.6	26.9	231.0	
Black	1,266	3	20	1,243	39.8	58.2	1.9	203.5	3	11	1,252	56.9	36.0	7.0	196.7	6	12	1,248	34.3	56.8	8.9	211.9	
Asian	1,189	0	10	1,179	12.9	74.5	12.6	224.4	1	5	1,183	17.6	43.3	39.1	232.9	3	3	1,183	12.4	52.6	35.0	234.0	
Pacific Islander	55	0	0	55	16.4	80.0	3.6	217.7	0	0	55	16.4	50.9	32.7	227.2	0	0	55	12.7	61.8	25.5	228.3	
Hispanic	1,042	2	9	1,031	36.5	62.5	1.1	203.9	3	5	1,034	49.7	41.2	9.1	201.3	8	5	1,029	31.9	57.6	10.5	214.1	
Amer Indian/AK Native	17	0	1	16	12.5	81.3	6.3	214.2	0	0	17	41.2	47.1	11.8	211.4	0	0	17	11.8	64.7	23.5	228.0	
<b>Economic Status</b>																							
Econ Disadvantaged	1,539	5	22	1,512	42.0	56.6	1.4	201.0	9	8	1,522	54.7	37.5	7.8	198.5	13	10	1,516	36.9	53.4	9.7	211.0	
Non-Econ Disadvantaged	12,185	16	104	12,065	17.8	76.4	5.8	218.2	25	59	12,101	28.0	48.1	23.9	220.2	56	51	12,078	13.4	60.1	26.5	230.4	

<sup>1</sup> Students appear in each applicable category, but they are included in Total Students only once.

<sup>2</sup> Excludes students who did not have Gender coded.

<sup>3</sup> Excludes students who did not have an ethnicity coded. Students with multiple ethnicities are reported for all ethnic categories that apply.



**New Jersey Statewide Testing System  
Grade Eight Proficiency Assessment  
Performance by Demographic Groups  
DFG GH**

Test Date: March 2004  
CYCLE II

	Language Arts Literacy						Mathematics						Science									
	Number Not Present	Number of Voids	Number of Valid Scale Scores	% Partially Proficient	% Proficient	Advanced Proficient	Scale Score Mean	Number Not Present	Number of Voids	Number of Valid Scale Scores	% Partially Proficient	% Proficient	Advanced Proficient	Scale Score Mean	Number Not Present	Number of Voids	Number of Valid Scale Scores	% Partially Proficient	% Proficient	Advanced Proficient	Scale Score Mean	
<b>Total Students<sup>1</sup></b>	27	99	14,502	15.6	76.4	8.0	221.0	31	65	14,532	25.4	46.4	28.2	223.6	50	58	14,520	14.4	56.9	28.7	231.2	
General Education	16	23	12,201	7.0	83.5	9.5	226.9	18	8	12,214	16.7	50.6	32.7	230.4	27	7	12,206	8.3	58.9	32.8	236.1	
Special Education	10	74	2,030	59.9	39.9	0.2	191.0	13	57	2,044	71.0	25.1	3.9	187.6	20	51	2,043	44.0	48.5	7.5	206.9	
Limited English Proficient	1	2	276	73.2	26.1	0.7	179.1	0	0	279	73.8	18.6	7.5	185.6	3	0	276	67.8	29.0	3.3	192.7	
<b>Gender<sup>2</sup></b>																						
Female	11	31	7,040	11.0	77.3	11.8	226.4	13	22	7,047	26.2	48.2	25.7	222.1	17	23	7,042	16.9	60.1	23.1	227.4	
Male	16	68	7,453	20.0	75.5	4.5	215.9	18	43	7,476	24.7	44.8	30.5	225.0	33	35	7,469	12.1	53.9	34.0	234.7	
<b>Migrant Status</b>																						
Migrant	2	0	2	0.0	100.0	0.0	209.0	0	0	2	100.0	0.0	0.0	178.0	0	0	2	50.0	50.0	0.0	207.0	
Non-Migrant	14,626	99	14,500	15.6	76.4	8.0	221.0	31	65	14,530	25.4	46.4	28.2	223.6	50	58	14,518	14.4	56.9	28.7	231.2	
<b>Ethnicity<sup>3</sup></b>																						
White	15	72	10,822	11.6	79.7	8.7	223.9	20	53	10,836	20.0	49.2	30.8	227.8	33	42	10,834	10.2	57.8	32.0	234.8	
Black	11	12	1,506	36.2	62.4	1.3	205.0	11	6	1,510	56.9	36.6	6.6	197.2	13	7	1,507	35.0	56.1	9.0	211.8	
Asian	0	5	1,063	9.2	74.2	16.6	229.2	0	4	1,064	13.2	38.3	48.5	238.9	0	6	1,062	8.9	50.2	41.0	239.8	
Pacific Islander	32	0	31	19.4	74.2	6.5	220.2	0	0	32	18.8	53.1	28.1	225.3	0	0	32	18.8	53.1	28.1	229.1	
Hispanic	0	6	1,047	32.4	65.1	2.5	206.5	0	2	1,051	47.0	41.0	12.0	204.2	3	3	1,047	32.3	56.8	10.9	213.7	
Amer. Indian/AK Native	20	0	20	25.0	75.0	0.0	210.2	0	0	20	30.0	40.0	30.0	215.8	0	0	20	25.0	40.0	35.0	225.1	
<b>Economic Status</b>																						
Econ Disadvantaged	8	9	1,309	42.3	56.4	1.3	200.8	8	6	1,312	56.6	34.9	8.5	197.3	10	8	1,308	41.0	50.4	8.6	208.5	
Non-Econ Disadvantaged	19	90	13,193	13.0	78.4	8.7	223.0	23	59	13,220	22.3	47.6	30.1	226.2	40	50	13,212	11.8	57.5	30.7	233.4	

<sup>1</sup> Students appear in each applicable category, but they are included in Total Students only once.

<sup>2</sup> Excludes students who did not have Gender coded.

<sup>3</sup> Excludes students who did not have an ethnicity coded. Students with multiple ethnicities are reported for all ethnic categories that apply.



## New Jersey Statewide Testing System Grade Eight Proficiency Assessment Performance by Demographic Groups DFG I

Test Date: March 2004  
CYCLE II

	Number Enrolled	Language Arts Literacy						Mathematics						Science									
		Number Not Present	Number of Voids	Number of Valid Scale Scores	% Partially Proficient	% Proficient	% Advanced Proficient	Scale Score Mean	Number Not Present	Number of Voids	Number of Valid Scale Scores	% Partially Proficient	% Proficient	% Advanced Proficient	Scale Score Mean	Number Not Present	Number of Voids	Number of Valid Scale Scores	% Partially Proficient	% Proficient	% Advanced Proficient	Scale Score Mean	
<b>Total Students<sup>1</sup></b>	20,548	40	142	20,366	10.3	77.8	11.9	226.5	50	98	20,400	17.7	46.1	36.3	231.2	64	80	20,404	8.8	53.8	37.5	237.9	
General Education	17,604	26	34	17,544	4.3	82.1	13.7	231.1	35	12	17,557	11.1	48.1	40.8	236.7	41	11	17,552	4.9	53.4	41.6	241.8	
Special Education	2,713	11	105	2,597	47.0	52.0	1.0	198.7	14	84	2,615	59.8	33.6	6.7	195.4	22	66	2,625	31.2	56.6	12.2	214.1	
Limited English Proficient	237	3	3	231	59.7	39.0	1.3	191.0	1	2	234	41.9	32.5	25.6	212.4	1	3	233	45.9	48.1	6.0	205.5	
<b>Gender<sup>2</sup></b>																							
Female	9,935	15	58	9,862	6.7	76.6	16.8	231.7	21	41	9,873	18.5	48.0	33.5	229.5	28	29	9,878	10.3	58.5	31.2	234.3	
Male	10,588	20	82	10,486	13.8	78.9	7.4	221.7	24	55	10,509	16.9	44.3	38.9	232.8	30	50	10,508	7.3	49.3	43.4	241.2	
<b>Migrant Status</b>																							
Migrant																							
Non-Migrant	20,548	40	142	20,366	10.3	77.8	11.9	226.5	50	98	20,400	17.7	46.1	36.3	231.2	64	80	20,404	8.8	53.8	37.5	237.9	
<b>Ethnicity<sup>3</sup></b>																							
White	16,281	24	107	16,150	9.0	79.5	11.5	227.3	33	72	16,176	16.0	47.8	36.2	231.9	38	59	16,184	7.5	53.9	38.6	239.0	
Black	948	3	11	934	30.7	67.1	2.1	209.1	3	8	937	48.9	42.7	8.4	202.8	4	8	936	27.2	62.9	9.8	215.9	
Asian	2,162	4	10	2,148	7.1	70.6	22.3	233.0	4	8	2,150	8.7	35.1	56.3	245.2	5	6	2,151	6.6	46.7	46.8	243.9	
Pacific Islander	106	1	0	105	5.7	74.3	20.0	231.7	2	0	104	11.5	36.6	53.8	242.2	2	0	104	8.7	40.4	51.0	244.4	
Hispanic	700	2	8	690	23.9	71.7	4.3	213.9	2	5	693	41.4	44.0	14.6	209.8	2	4	694	21.5	61.4	17.1	222.1	
Amer Indian/AK Native	21	0	1	20	10.0	75.0	15.0	227.1	0	1	20	20.0	45.0	35.0	225.6	0	1	20	5.0	65.0	30.0	238.1	
<b>Economic Status</b>																							
Econ Disadvantaged	731	4	15	712	37.4	60.5	2.1	205.1	3	13	715	48.4	41.5	10.1	203.1	4	11	716	32.0	58.1	9.9	213.5	
Non-Econ Disadvantaged	19,817	36	127	19,654	9.3	78.4	12.3	227.3	47	85	19,685	16.5	46.3	37.2	232.2	60	69	19,688	7.9	53.6	38.5	238.8	

<sup>1</sup> Students appear in each applicable category, but they are included in Total Students only once.

<sup>2</sup> Excludes students who did not have Gender coded.

<sup>3</sup> Excludes students who did not have an ethnicity coded. Students with multiple ethnicities are reported for all ethnic categories that apply.



**New Jersey Statewide Testing System  
Grade Eight Proficiency Assessment  
Performance by Demographic Groups  
DFG J**

Test Date: March 2004  
CYCLE II

	Language Arts Literacy						Mathematics						Science						
	Number Enrolled	Number Not Present	Number of Voids	Number of Valid Scale Scores	% Partially Proficient	% Advanced Proficient	Scale Score Mean	Number Not Present	Number of Voids	Number of Valid Scale Scores	% Partially Proficient	% Advanced Proficient	Scale Score Mean	Number Not Present	Number of Voids	Number of Valid Scale Scores	% Partially Proficient	% Advanced Proficient	Scale Score Mean
<b>Total Students<sup>1</sup></b>	1,623	1	9	1,613	7.0	76.8	231.1	2	6	1,615	12.3	43.5	236.8	2	7	1,614	5.7	47.9	243.5
General Education	1,405	1	6	1,398	2.2	79.1	235.5	1	3	1,401	7.1	43.4	242.0	2	3	1,400	3.2	46.1	246.9
Special Education	201	0	3	198	36.4	63.6	203.6	1	3	197	47.2	45.7	200.8	0	4	197	21.8	59.4	221.7
Limited English Proficient	17	0	0	17	58.8	41.2	190.8	0	0	17	29.4	29.4	228.6	0	0	17	23.5	64.7	217.7
<b>Gender<sup>2</sup></b>																			
Female	801	1	5	795	3.8	71.9	237.0	1	3	797	12.9	44.0	235.9	2	3	796	6.5	52.5	240.4
Male	821	0	4	817	10.0	81.6	225.5	1	3	817	11.5	43.1	237.8	0	4	817	4.8	43.5	246.6
<b>Migrant Status</b>																			
Migrant																			
Non-Migrant	1,623	1	9	1,613	7.0	76.8	231.1	2	6	1,615	12.3	43.5	236.8	2	7	1,614	5.7	47.9	243.5
<b>Ethnicity<sup>3</sup></b>																			
White	1,395	1	8	1,386	6.4	78.1	231.1	2	5	1,388	12.1	45.4	235.8	2	6	1,387	5.3	49.2	243.1
Black	14	0	0	14	14.3	85.7	219.6	0	0	14	50.0	35.7	206.0	0	0	14	14.3	57.1	230.8
Asian	173	0	1	172	5.2	69.2	236.2	0	1	172	5.2	29.1	252.3	0	1	172	3.5	37.2	251.5
Pacific Islander	3	0	0	3	0.0	100.0	226.3	0	0	3	33.3	66.7	217.7	0	0	3	0.0	100.0	239.7
Hispanic	37	0	0	37	35.1	59.5	211.4	0	0	37	35.1	45.9	212.5	0	0	37	27.0	45.9	224.0
Amer Indian/AK Native																			
<b>Economic Status</b>																			
Econ Disadvantaged	9	0	1	8	62.5	37.5	187.5	0	0	9	77.8	11.1	185.6	1	0	8	37.5	37.5	214.4
Non-Econ Disadvantaged	1,614	1	8	1,605	6.7	77.0	231.3	2	6	1,606	11.9	43.7	237.1	1	7	1,606	5.5	47.9	243.7

<sup>1</sup> Students appear in each applicable category, but they are included in Total Students only once.

<sup>2</sup> Excludes students who did not have gender coded.

<sup>3</sup> Excludes students who did not have an ethnicity coded. Students with multiple ethnicities are reported for all ethnic categories that apply.



**New Jersey Statewide Testing System  
Grade Eight Proficiency Assessment  
Performance by Demographic Groups  
DFG 0**

Test Date: March 2004  
CYCLE II

	Number Enrolled	Language Arts Literacy				Mathematics				Science												
		Number Not Present <sup>1</sup>	Number of Valid Scores	% Partially Proficient	% Proficient	Advanced Proficient	Scale Score Mean	Number Not Present	Number of Valid Scores	% Partially Proficient	% Proficient	Advanced Proficient	Scale Score Mean									
<b>Total Students<sup>1</sup></b>	174	13	146	95.9	4.1	0.0	157.6	12	2	160	98.1	1.9	0.0	161.6	7	4	163	81.6	18.4	0.0	181.6	
General Education	61	4	53	96.2	3.8	0.0	164.5	2	0	59	96.6	3.4	0.0	164.0	0	0	61	82.0	18.0	0.0	183.6	
Special Education	113	9	93	95.7	4.3	0.0	153.7	10	2	101	99.0	1.0	0.0	160.2	7	4	102	81.4	18.6	0.0	180.5	
Limited English Proficient																						
<b>Gender<sup>2</sup></b>	49	5	39	89.7	10.3	0.0	164.9	7	1	41	92.7	7.3	0.0	164.5	5	1	43	76.7	23.3	0.0	181.0	
Female	123	8	105	98.1	1.9	0.0	155.2	5	1	117	100.0	0.0	0.0	160.5	2	3	118	83.1	16.9	0.0	181.9	
Male																						
<b>Migrant Status</b>	174	13	146	95.9	4.1	0.0	157.6	12	2	160	98.1	1.9	0.0	161.6	7	4	163	81.6	18.4	0.0	181.6	
Migrant																						
Non-Migrant																						
<b>Ethnicity<sup>3</sup></b>	37	2	30	90.0	10.0	0.0	156.8	1	2	34	97.1	2.9	0.0	163.2	1	2	34	67.6	32.4	0.0	185.4	
White	118	11	99	97.0	3.0	0.0	158.0	10	0	108	98.1	1.9	0.0	160.9	5	2	111	84.7	15.3	0.0	180.4	
Black	1	0	1	100.0	0.0	0.0	152.0	0	0	1	100.0	0.0	0.0	179.0	0	0	1	100.0	0.0	0.0	182.0	
Asian																						
Pacific Islander																						
Hispanic	27	1	24	100.0	0.0	0.0	154.4	2	0	25	100.0	0.0	0.0	158.9	1	0	26	88.5	11.5	0.0	181.7	
Amer Indian/AK Native																						
<b>Economic Status</b>	130	13	104	94.2	5.8	0.0	159.2	10	2	118	97.5	2.5	0.0	162.0	6	4	120	80.8	19.2	0.0	182.3	
Econ Disadvantaged	44	0	42	100.0	0.0	0.0	153.8	2	0	42	100.0	0.0	0.0	160.4	1	0	43	83.7	16.3	0.0	179.6	
Non-Econ Disadvantaged																						

<sup>1</sup> Students appear in each applicable category, but they are included in Total Students only once.

<sup>2</sup> Excludes students who did not have Gender coded.

<sup>3</sup> Excludes students who did not have an ethnicity coded. Students with multiple ethnicities are reported for all ethnic categories that apply.



## New Jersey Statewide Testing System Grade Eight Proficiency Assessment Performance by Demographic Groups DFG R

Test Date: March 2004  
CYCLE II

	Language Arts Literacy						Mathematics						Science										
	Number Not Present	Number of Valid Scale Scores	% Partially Proficient	% Advanced Proficient	Scale Score Mean	Number Not Present	Number of Valid Scale Scores	% Partially Proficient	% Advanced Proficient	Scale Score Mean	Number Not Present	Number of Valid Scale Scores	% Partially Proficient	% Advanced Proficient	Scale Score Mean	Number Not Present	Number of Valid Scale Scores	% Partially Proficient	% Advanced Proficient	Scale Score Mean			
<b>Total Students<sup>1</sup></b>	993																						
General Education	0	898	46.8	52.0	199.5	1	904	64.5	27.4	193.1	4	901	64.4	10.1	208.1	4	901	64.4	10.1	208.1	4	901	
Special Education	2	85	82.4	17.6	172.5	2	85	83.5	12.9	178.0	2	85	70.6	23.5	194.2	2	85	70.6	23.5	194.2	2	85	
Limited English Proficient	3	3	100.0	0.0	178.7	0	3	100.0	0.0	171.3	0	3	66.7	33.3	188.0	0	3	66.7	33.3	188.0	0	3	
<b>Gender<sup>2</sup></b>																							
Female	0	505	39.2	59.4	203.8	0	509	67.4	25.5	191.7	1	508	46.3	45.7	205.5	1	508	46.3	45.7	205.5	1	508	
Male	2	477	61.0	38.2	190.2	3	479	64.7	26.9	192.1	5	477	47.0	11.5	208.6	5	477	47.0	11.5	208.6	5	477	
<b>Migrant Status</b>																							
Migrant	2	984	49.9	49.0	197.1	3	990	66.2	26.2	191.8	6	987	46.7	43.6	206.9	6	987	46.7	43.6	206.9	6	987	
Non-Migrant	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
<b>Ethnicity<sup>3</sup></b>																							
White	0	129	14.0	79.1	219.5	0	129	25.6	44.2	224.9	0	129	10.1	50.4	237.8	0	129	10.1	50.4	237.8	0	129	
Black	2	637	61.9	38.1	189.9	3	643	78.1	19.8	182.2	6	640	57.5	40.0	197.9	6	640	57.5	40.0	197.9	6	640	
Asian	0	26	15.4	80.8	220.3	0	26	30.8	26.9	224.0	0	26	15.4	34.6	241.3	0	26	15.4	34.6	241.3	0	26	
Pacific Islander	0	1	0.0	100.0	217.0	0	1	100.0	0.0	194.0	0	1	0.0	100.0	204.0	0	1	0.0	100.0	204.0	0	1	
Hispanic	0	173	36.4	63.6	204.2	0	173	54.9	36.2	199.3	0	173	35.3	56.1	213.7	0	173	35.3	56.1	213.7	0	173	
Amer. Indian/AK Native	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
<b>Economic Status</b>																							
Econ Disadvantaged	1	570	59.3	40.5	190.9	2	575	73.4	23.5	185.1	5	572	54.9	40.9	199.9	5	572	54.9	40.9	199.9	5	572	
Non-Econ Disadvantaged	1	414	37.0	60.6	205.7	1	415	56.1	29.9	201.1	1	415	35.4	17.3	216.4	1	415	35.4	17.3	216.4	1	415	

<sup>1</sup> Students appear in each applicable category, but they are included in Total Students only once.

<sup>2</sup> Excludes students who did not have Gender coded.

<sup>3</sup> Excludes students who did not have an ethnicity coded. Students with multiple ethnicities are reported for all ethnic categories that apply.



## New Jersey Statewide Testing System Grade Eight Proficiency Assessment Performance by Demographic Groups Special Needs

Test Date: March 2004  
CYCLE II

	Language Arts Literacy										Mathematics										Science									
	Number Enrolled	Number Not Present	Number of Voids	Number of Valid Scale Scores	% Partially Proficient	% Proficient	% Advanced Proficient	Scale Score Mean	Number Not Present	Number of Voids	Number of Valid Scale Scores	% Partially Proficient	% Proficient	% Advanced Proficient	Scale Score Mean	Number Not Present	Number of Voids	Number of Valid Scale Scores	% Partially Proficient	% Proficient	% Advanced Proficient	Scale Score Mean								
<b>Total Students<sup>1</sup></b>	22,150	251	471	21,428	55.4	43.6	1.0	192.6	319	170	21,661	65.5	28.6	5.9	190.5	376	149	21,625	54.2	41.3	4.5	199.9								
General Education	15,640	120	171	15,349	42.2	56.4	1.4	202.9	160	28	15,452	56.4	35.8	7.9	198.0	177	24	15,439	43.8	50.2	6.0	206.0								
Special Education	4,458	107	279	4,072	90.9	9.1	0.0	165.5	131	141	4,186	91.8	7.7	0.5	168.5	172	119	4,167	81.4	18.0	0.6	183.9								
Limited English Proficient	2,195	25	32	2,138	85.4	14.6	0.0	168.0	29	9	2,157	80.6	17.3	2.1	178.2	31	12	2,152	78.4	20.9	0.7	185.7								
<b>Gender<sup>2</sup></b>																														
Female	10,902	106	165	10,631	47.4	50.9	1.7	198.8	132	61	10,709	64.8	29.7	5.5	191.1	151	50	10,701	56.7	39.8	3.4	198.2								
Male	11,215	142	303	10,770	63.4	36.3	0.4	186.4	183	108	10,924	66.1	27.7	6.3	189.9	221	97	10,897	51.8	42.7	5.5	201.6								
<b>Migrant Status</b>																														
Migrant	29	1	1	27	74.1	25.9	0.0	180.6	1	1	27	81.5	14.8	3.7	180.5	2	0	27	55.6	40.7	3.7	193.1								
Non-Migrant	22,121	250	470	21,401	55.4	43.6	1.0	192.6	318	169	21,634	65.4	28.7	5.9	190.5	374	149	21,598	54.2	41.3	4.5	199.9								
<b>Ethnicity<sup>3</sup></b>																														
White	2,806	20	48	2,738	35.1	62.3	2.6	206.0	24	20	2,762	44.8	42.6	12.6	206.1	30	23	2,753	30.6	55.8	13.6	216.1								
Black	9,508	133	239	9,136	62.7	36.9	0.5	188.7	176	76	9,256	75.8	21.5	2.6	183.0	212	66	9,230	63.7	34.4	1.9	194.1								
Asian	526	2	3	521	33.8	62.6	3.6	206.6	3	3	520	39.0	41.0	20.0	212.5	3	3	520	33.8	50.8	15.4	215.0								
Pacific Islander	29	0	0	29	34.5	65.5	0.0	205.8	0	0	29	41.4	37.9	20.7	211.8	0	0	29	34.5	48.3	17.2	215.1								
Hispanic	9,212	94	176	8,942	55.3	43.7	0.9	191.8	112	70	9,030	62.6	31.1	6.3	192.3	127	55	9,030	52.9	43.4	3.7	200.1								
Amer Indian/AK Native	50	0	0	50	50.0	50.0	0.0	196.8	0	0	50	56.0	30.0	14.0	197.6	0	0	50	48.0	42.0	10.0	205.2								
<b>Economic Status</b>																														
Econ Disadvantaged	15,723	168	322	15,233	58.4	40.8	0.7	190.7	218	111	15,394	67.6	27.3	5.1	188.8	260	97	15,366	57.5	39.1	3.4	197.6								
Non-Econ Disadvantaged	6,427	83	149	6,195	48.1	50.3	1.7	197.1	101	59	6,267	60.2	31.8	8.0	194.6	116	52	6,259	46.3	46.5	7.2	205.3								

<sup>1</sup> Students appear in each applicable category, but they are included in Total Students only once.

<sup>2</sup> Excludes students who did not have Gender coded.

<sup>3</sup> Excludes students who did not have an ethnicity coded. Students with multiple ethnicities are reported for all ethnic categories that apply.



## New Jersey Statewide Testing System Grade Eight Proficiency Assessment Performance by Demographic Groups Non Special Needs

Test Date: March 2004  
CYCLE II

	Number Enrolled	Language Arts Literacy					Mathematics					Science				
		Number Not Present	Number of Valid Scores	% Partially Proficient	% Proficient	Scale Score Mean	Number Not Present	Number of Valid Scores	% Partially Proficient	% Proficient	Scale Score Mean	Number Not Present	Number of Valid Scores	% Partially Proficient	% Proficient	Scale Score Mean
<b>Total Students<sup>1</sup></b>	<b>88,120</b>	<b>293</b>	<b>86,999</b>	<b>21.6</b>	<b>71.9</b>	<b>6.6</b>	<b>368</b>	<b>87,304</b>	<b>31.5</b>	<b>44.9</b>	<b>23.5</b>	<b>510</b>	<b>87,216</b>	<b>18.1</b>	<b>56.7</b>	<b>227.7</b>
General Education	72,840	157	72,430	12.3	79.9	7.8	193	72,593	22.8	49.6	27.6	269	72,512	11.6	59.2	232.8
Special Education	13,552	123	12,871	66.5	33.2	0.3	165	12,998	75.1	21.9	3.0	226	12,997	47.7	46.2	203.9
Limited English Proficient	1,787	14	1,755	79.0	20.6	0.4	11	1,771	71.4	20.9	7.6	16	1,764	68.5	29.4	191.3
<b>Gender<sup>2</sup></b>																
Female	42,702	112	42,286	15.6	74.9	9.5	157	42,384	32.2	46.1	21.6	211	42,341	20.4	59.2	224.4
Male	45,327	173	44,636	27.2	69.0	3.8	203	44,842	30.9	43.8	25.4	290	44,797	15.9	54.3	230.9
<b>Migrant Status</b>																
Migrant	45	0	43	51.2	44.2	4.7	0	44	68.2	25.0	6.8	1	44	52.3	45.5	202.7
Non-Migrant	88,075	293	86,956	21.6	71.9	6.6	368	87,260	31.5	44.9	23.5	509	87,172	18.1	56.7	227.7
<b>Ethnicity<sup>3</sup></b>																
White	61,891	145	61,255	16.0	76.7	7.3	203	61,401	24.9	48.5	26.6	273	61,374	12.4	58.3	232.4
Black	10,250	82	9,988	45.2	53.7	1.2	95	10,069	64.1	31.3	4.5	124	10,046	41.5	51.9	207.3
Asian	6,223	6	6,178	12.6	72.0	15.4	5	6,191	15.2	40.3	44.5	10	6,188	11.4	50.8	236.9
Pacific Islander	443	1	439	16.4	75.4	8.2	2	441	20.4	45.6	34.0	2	441	13.4	58.3	231.6
Hispanic	8,723	45	8,589	40.0	58.3	1.7	51	8,632	52.4	39.0	8.5	75	8,610	35.4	55.5	211.6
Amer Indian/AK Native	122	0	118	22.9	72.9	4.2	0	120	35.0	45.8	19.2	0	120	20.8	58.3	223.2
<b>Economic Status</b>																
Econ Disadvantaged	13,577	113	13,231	47.7	51.2	1.1	129	13,350	59.5	33.6	6.9	167	13,305	41.2	51.1	208.0
Non-Econ Disadvantaged	74,543	180	73,768	16.9	75.6	7.5	239	73,954	26.5	47.0	26.5	343	73,911	13.9	57.7	231.2

<sup>1</sup> Students appear in each applicable category, but they are included in Total Students only once.

<sup>2</sup> Excludes students who did not have gender coded.

<sup>3</sup> Excludes students who did not have an ethnicity coded. Students with multiple ethnicities are reported for all ethnic categories that apply.



# **Appendix C**

## **Raw to Scale Scores Conversions**

## 2004 GEPA LAL Raw Score to Scale Score

RS	Ability	SS
0.0	-5.2310	103
0.5	-4.8875	105
1.0	-4.5440	106
1.5	-4.2015	108
2.0	-3.8590	110
2.5	-3.6585	112
3.0	-3.4580	114
3.5	-3.3145	115
4.0	-3.1710	117
4.5	-3.0580	119
5.0	-2.9450	121
5.5	-2.8505	122
6.0	-2.7560	124
6.5	-2.6725	126
7.0	-2.5890	127
7.5	-2.5125	129
8.0	-2.4360	131
8.5	-2.3650	133
9.0	-2.2940	135
9.5	-2.2260	136
10.0	-2.1580	138
10.5	-2.0930	140
11.0	-2.0280	142
11.5	-1.9640	143
12.0	-1.9000	145
12.5	-1.8380	147
13.0	-1.7760	148
13.5	-1.7150	150
14.0	-1.6540	152
14.5	-1.5945	153
15.0	-1.5350	154
15.5	-1.4765	156
16.0	-1.4180	157
16.5	-1.3610	159
17.0	-1.3040	160
17.5	-1.2475	161
18.0	-1.1910	163
18.5	-1.1355	164
19.0	-1.0800	165
19.5	-1.0250	167
20.0	-0.9700	168
20.5	-0.9160	169
21.0	-0.8620	171

RS	Ability	SS
21.5	-0.8080	172
22.0	-0.7540	174
22.5	-0.7005	175
23.0	-0.6470	176
23.5	-0.5930	178
24.0	-0.5390	179
24.5	-0.4855	180
25.0	-0.4320	182
25.5	-0.3775	183
26.0	-0.3230	185
26.5	-0.2685	186
27.0	-0.2140	187
27.5	-0.1585	189
28.0	-0.1030	190
28.5	-0.0460	192
29.0	0.0110	193
29.5	0.0685	195
30.0	0.1260	196
30.5	0.1850	198
31.0	0.2440	200
31.5	0.3040	201
32.0	0.3640	202
32.5	0.4275	204
33.0	0.4910	206
33.5	0.5555	207
34.0	0.6200	209
34.5	0.6865	211
35.0	0.7530	212
35.5	0.8225	214
36.0	0.8920	215
36.5	0.9645	217
37.0	1.0370	219
37.5	1.1120	221
38.0	1.1870	222
38.5	1.2660	224
39.0	1.3450	226
39.5	1.4275	227
40.0	1.5100	229
40.5	1.5965	231
41.0	1.6830	232
41.5	1.7730	234
42.0	1.8630	236
42.5	1.9580	237

RS	Ability	SS
43.0	2.0530	239
43.5	2.1515	241
44.0	2.2500	242
44.5	2.3525	244
45.0	2.4550	245
45.5	2.5600	247
46.0	2.6650	249
46.5	2.7730	250
47.0	2.8810	252
47.5	2.9880	254
48.0	3.0950	255
48.5	3.2050	257
49.0	3.3150	259
49.5	3.4235	260
50.0	3.5320	262
50.5	3.6405	264
51.0	3.7490	265
51.5	3.8585	267
52.0	3.9680	269
52.5	4.0805	271
53.0	4.1930	272
53.5	4.3100	274
54.0	4.4270	276
54.5	4.5520	278
55.0	4.6770	279
55.5	4.8135	281
56.0	4.9500	283
56.5	5.1020	285
57.0	5.2540	287
57.5	5.4280	289
58.0	5.6020	290
58.5	5.8095	292
59.0	6.0170	294
59.5	6.2840	296
60.0	6.5510	298
60.5	6.9610	300
61.0	7.3710	300
61.5	7.7485	300
62.0	8.1260	300

## 2004 GEPA Mathematics Raw Score to Scale Score

RS	Ability	SS
0.0	-5.0150	137
0.5	-4.6550	137
1.0	-4.2950	139
1.5	-3.9225	140
2.0	-3.5500	141
2.5	-3.3220	142
3.0	-3.0940	143
3.5	-2.9255	144
4.0	-2.7570	146
4.5	-2.6220	147
5.0	-2.4870	148
5.5	-2.3725	149
6.0	-2.2580	150
6.5	-2.1580	152
7.0	-2.0580	153
7.5	-1.9690	154
8.0	-1.8800	155
8.5	-1.7985	156
9.0	-1.7170	158
9.5	-1.6425	159
10.0	-1.5680	160
10.5	-1.4985	162
11.0	-1.4290	163
11.5	-1.3635	164
12.0	-1.2980	166
12.5	-1.2365	167
13.0	-1.1750	168
13.5	-1.1165	169
14.0	-1.0580	171
14.5	-1.0015	172
15.0	-0.9450	174
15.5	-0.8915	175
16.0	-0.8380	176

RS	Ability	SS
16.5	-0.7860	178
17.0	-0.7340	179
17.5	-0.6835	180
18.0	-0.6330	182
18.5	-0.5845	183
19.0	-0.5360	185
19.5	-0.4885	186
20.0	-0.4410	188
20.5	-0.3945	189
21.0	-0.3480	191
21.5	-0.3025	192
22.0	-0.2570	194
22.5	-0.2120	195
23.0	-0.1670	197
23.5	-0.1230	198
24.0	-0.0790	200
24.5	-0.0355	201
25.0	0.0080	203
25.5	0.0515	204
26.0	0.0950	206
26.5	0.1385	208
27.0	0.1820	209
27.5	0.2255	211
28.0	0.2690	212
28.5	0.3130	214
29.0	0.3570	216
29.5	0.4015	218
30.0	0.4460	219
30.5	0.4915	221
31.0	0.5370	223
31.5	0.5835	224
32.0	0.6300	226
32.5	0.6780	228

RS	Ability	SS
33.0	0.7260	230
33.5	0.7755	232
34.0	0.8250	234
34.5	0.8770	235
35.0	0.9290	237
35.5	0.9835	239
36.0	1.0380	241
36.5	1.0960	243
37.0	1.1540	245
37.5	1.2160	247
38.0	1.2780	250
38.5	1.3450	251
39.0	1.4120	252
39.5	1.4855	254
40.0	1.5590	256
40.5	1.6410	258
41.0	1.7230	260
41.5	1.8155	262
42.0	1.9080	264
42.5	2.0160	266
43.0	2.1240	267
43.5	2.2535	269
44.0	2.3830	271
44.5	2.5460	273
45.0	2.7090	275
45.5	2.9330	276
46.0	3.1570	278
46.5	3.5270	280
47.0	3.8970	281
47.5	4.2565	282
48.0	4.6160	284

## 2004 GEPA Science Raw Score to Scale Score

RS	Ability	SS
0.0	-4.549	126
0.5	-4.195	128
1.0	-3.841	129
1.5	-3.480	131
2.0	-3.118	133
2.5	-2.901	135
3.0	-2.683	137
3.5	-2.524	139
4.0	-2.365	141
4.5	-2.239	143
5.0	-2.112	145
5.5	-2.006	146
6.0	-1.899	148
6.5	-1.806	150
7.0	-1.713	152
7.5	-1.631	154
8.0	-1.549	156
8.5	-1.474	158
9.0	-1.399	159
9.5	-1.330	161
10.0	-1.261	163
10.5	-1.197	165
11.0	-1.134	166
11.5	-1.074	168
12.0	-1.014	170
12.5	-0.957	172
13.0	-0.900	173
13.5	-0.846	175
14.0	-0.792	177
14.5	-0.740	179
15.0	-0.689	180
15.5	-0.639	182
16.0	-0.589	184
16.5	-0.541	185
17.0	-0.493	188
17.5	-0.447	189

RS	Ability	SS
18.0	-0.400	191
18.5	-0.354	192
19.0	-0.309	194
19.5	-0.264	195
20.0	-0.220	197
20.5	-0.176	198
21.0	-0.132	200
21.5	-0.090	202
22.0	-0.047	204
22.5	-0.004	205
23.0	0.038	207
23.5	0.080	208
24.0	0.122	210
24.5	0.163	211
25.0	0.205	213
25.5	0.247	214
26.0	0.288	216
26.5	0.329	217
27.0	0.371	219
27.5	0.412	220
28.0	0.454	222
28.5	0.495	223
29.0	0.537	225
29.5	0.579	227
30.0	0.621	228
30.5	0.664	230
31.0	0.706	231
31.5	0.749	232
32.0	0.792	234
32.5	0.836	235
33.0	0.880	237
33.5	0.924	238
34.0	0.969	239
34.5	1.015	241
35.0	1.061	242
35.5	1.108	244

RS	Ability	SS
36.0	1.155	245
36.5	1.203	246
37.0	1.251	248
37.5	1.301	249
38.0	1.352	250
38.5	1.405	252
39.0	1.457	253
39.5	1.512	254
40.0	1.566	256
40.5	1.623	257
41.0	1.681	258
41.5	1.742	259
42.0	1.803	261
42.5	1.868	263
43.0	1.933	264
43.5	2.003	265
44.0	2.072	266
44.5	2.148	267
45.0	2.223	269
45.5	2.306	270
46.0	2.388	271
46.5	2.479	272
47.0	2.570	273
47.5	2.673	274
48.0	2.776	276
48.5	2.894	277
49.0	3.011	278
49.5	3.150	279
50.0	3.289	280
50.5	3.461	281
51.0	3.632	282
51.5	3.862	283
52.0	4.091	284
52.5	4.462	286
53.0	4.832	288
53.5	5.189	291
54.0	5.546	294

## 2004 GEPA LAL Braille Raw Score to Scale Score

RS	Ability	SS
0.0	-4.8880	105
0.5	-4.5490	106
1.0	-4.2100	108
1.5	-3.8785	110
2.0	-3.5470	113
2.5	-3.3570	115
3.0	-3.1670	117
3.5	-3.0325	119
4.0	-2.8980	121
4.5	-2.7905	123
5.0	-2.6830	125
5.5	-2.5905	127
6.0	-2.4980	130
6.5	-2.4150	131
7.0	-2.3320	134
7.5	-2.2545	136
8.0	-2.1770	138
8.5	-2.1035	140
9.0	-2.0300	142
9.5	-1.9590	144
10.0	-1.8880	145
10.5	-1.8200	147
11.0	-1.7520	149
11.5	-1.6855	151
12.0	-1.6190	152
12.5	-1.5545	154
13.0	-1.4900	156
13.5	-1.4270	157
14.0	-1.3640	159
14.5	-1.3030	160
15.0	-1.2420	162
15.5	-1.1820	163
16.0	-1.1220	164
16.5	-1.0630	166
17.0	-1.0040	167
17.5	-0.9460	169
18.0	-0.8880	170
18.5	-0.8310	172

RS	Ability	SS
19.0	-0.7740	173
19.5	-0.7170	174
20.0	-0.6600	176
20.5	-0.6030	177
21.0	-0.5460	179
21.5	-0.4895	180
22.0	-0.4330	182
22.5	-0.3755	183
23.0	-0.3180	185
23.5	-0.2605	186
24.0	-0.2030	188
24.5	-0.1440	189
25.0	-0.0850	191
25.5	-0.0255	192
26.0	0.0340	194
26.5	0.0950	195
27.0	0.1560	197
27.5	0.2190	199
28.0	0.2820	200
28.5	0.3465	202
29.0	0.4110	204
29.5	0.4770	205
30.0	0.5430	207
30.5	0.6130	209
31.0	0.6830	210
31.5	0.7550	212
32.0	0.8270	214
32.5	0.9015	216
33.0	0.9760	217
33.5	1.0545	219
34.0	1.1330	221
34.5	1.2155	223
35.0	1.2980	225
35.5	1.3840	226
36.0	1.4700	228
36.5	1.5610	230
37.0	1.6520	232
37.5	1.7475	233

RS	Ability	SS
38.0	1.8430	235
38.5	1.9430	237
39.0	2.0430	239
39.5	2.1485	240
40.0	2.2540	242
40.5	2.3635	244
41.0	2.4730	246
41.5	2.5860	247
42.0	2.6990	250
42.5	2.8155	251
43.0	2.9320	253
43.5	3.0500	255
44.0	3.1680	256
44.5	3.2855	258
45.0	3.4030	260
45.5	3.5230	262
46.0	3.6430	264
46.5	3.7630	266
47.0	3.8830	268
47.5	4.0065	269
48.0	4.1300	271
48.5	4.2600	273
49.0	4.3900	275
49.5	4.5310	277
50.0	4.6720	279
50.5	4.8295	281
51.0	4.9870	283
51.5	5.1680	286
52.0	5.3490	288
52.5	5.5665	290
53.0	5.7840	292
53.5	6.0630	294
54.0	6.3420	296
54.5	6.7655	299
55.0	7.1890	300
55.5	7.5735	300
56.0	7.9580	300

## 2004 GEPA Mathematics Braille Raw Score to Scale Score

RS	Ability	SS
0.0	-4.8690	137
0.5	-4.5045	138
1.0	-4.1400	139
1.5	-3.7595	140
2.0	-3.3790	142
2.5	-3.1435	143
3.0	-2.9080	145
3.5	-2.7325	146
4.0	-2.5570	147
4.5	-2.4145	149
5.0	-2.2720	150
5.5	-2.1510	152
6.0	-2.0300	153
6.5	-1.9235	155
7.0	-1.8170	156
7.5	-1.7215	158
8.0	-1.6260	159
8.5	-1.5385	161
9.0	-1.4510	162
9.5	-1.3700	164
10.0	-1.2890	166
10.5	-1.2135	167
11.0	-1.1380	169
11.5	-1.0670	171
12.0	-0.9960	172
12.5	-0.9285	174
13.0	-0.8610	176
13.5	-0.7970	177

RS	Ability	SS
14.0	-0.7330	179
14.5	-0.6710	181
15.0	-0.6090	183
15.5	-0.5495	184
16.0	-0.4900	186
16.5	-0.4325	188
17.0	-0.3750	190
17.5	-0.3190	192
18.0	-0.2630	193
18.5	-0.2080	195
19.0	-0.1530	197
19.5	-0.0990	200
20.0	-0.0450	201
20.5	0.0085	203
21.0	0.0620	205
21.5	0.1155	207
22.0	0.1690	209
22.5	0.2230	211
23.0	0.2770	213
23.5	0.3315	215
24.0	0.3860	217
24.5	0.4415	219
25.0	0.4970	221
25.5	0.5545	223
26.0	0.6120	226
26.5	0.6725	228
27.0	0.7330	230
27.5	0.7960	232

RS	Ability	SS
28.0	0.8590	235
28.5	0.9270	237
29.0	0.9950	240
29.5	1.0680	242
30.0	1.1410	244
30.5	1.2220	247
31.0	1.3030	250
31.5	1.3940	252
32.0	1.4850	254
32.5	1.5910	257
33.0	1.6970	259
33.5	1.8235	262
34.0	1.9500	265
34.5	2.1105	267
35.0	2.2710	270
35.5	2.4920	272
36.0	2.7130	275
36.5	3.0805	277
37.0	3.4480	279
37.5	3.8060	281
38.0	4.1640	282

## 2004 GEPA Science Braille Raw Score to Scale Score

RS	Ability	SS
0.0	-4.509	127
0.5	-4.155	128
1.0	-3.800	130
1.5	-3.438	131
2.0	-3.076	133
2.5	-2.857	135
3.0	-2.638	138
3.5	-2.478	140
4.0	-2.318	142
4.5	-2.191	143
5.0	-2.063	145
5.5	-1.956	147
6.0	-1.848	149
6.5	-1.754	151
7.0	-1.660	153
7.5	-1.577	155
8.0	-1.493	157
8.5	-1.417	159
9.0	-1.341	161
9.5	-1.271	163
10.0	-1.201	165
10.5	-1.136	166
11.0	-1.071	168
11.5	-1.010	170
12.0	-0.948	172
12.5	-0.890	174
13.0	-0.832	176
13.5	-0.776	178
14.0	-0.721	179
14.5	-0.668	181
15.0	-0.615	183
15.5	-0.563	184
16.0	-0.512	187
16.5	-0.463	188

RS	Ability	SS
17.0	-0.413	190
17.5	-0.364	192
18.0	-0.316	194
18.5	-0.269	195
19.0	-0.222	197
19.5	-0.176	198
20.0	-0.129	200
20.5	-0.083	202
21.0	-0.038	204
21.5	0.007	206
22.0	0.052	207
22.5	0.097	209
23.0	0.141	210
23.5	0.185	212
24.0	0.229	214
24.5	0.273	216
25.0	0.317	217
25.5	0.361	219
26.0	0.405	220
26.5	0.449	222
27.0	0.494	223
27.5	0.539	225
28.0	0.583	227
28.5	0.628	228
29.0	0.673	230
29.5	0.719	231
30.0	0.764	233
30.5	0.811	234
31.0	0.857	236
31.5	0.905	238
32.0	0.952	239
32.5	1.000	241
33.0	1.049	242
33.5	1.099	243

RS	Ability	SS
34.0	1.149	245
34.5	1.201	246
35.0	1.253	248
35.5	1.306	250
36.0	1.360	251
36.5	1.416	252
37.0	1.473	253
37.5	1.532	255
38.0	1.591	256
38.5	1.653	258
39.0	1.715	259
39.5	1.781	260
40.0	1.848	262
40.5	1.919	263
41.0	1.990	265
41.5	2.067	266
42.0	2.144	267
42.5	2.228	269
43.0	2.313	270
43.5	2.406	271
44.0	2.499	272
44.5	2.603	273
45.0	2.708	275
45.5	2.828	276
46.0	2.948	278
46.5	3.089	279
47.0	3.231	280
47.5	3.405	281
48.0	3.579	282
48.5	3.811	283
49.0	4.043	284
49.5	4.416	285
50.0	4.789	288
50.5	5.147	291
51.0	5.504	294

## 2004 GEPA Science Large Print Raw Score to Scale Score

RS	Ability	SS
0.0	-4.541	126
0.5	-4.187	128
1.0	-3.833	129
1.5	-3.471	131
2.0	-3.110	133
2.5	-2.892	135
3.0	-2.674	137
3.5	-2.515	139
4.0	-2.356	141
4.5	-2.229	143
5.0	-2.102	145
5.5	-1.995	147
6.0	-1.888	148
6.5	-1.796	151
7.0	-1.703	152
7.5	-1.620	154
8.0	-1.537	156
8.5	-1.462	158
9.0	-1.387	160
9.5	-1.318	162
10.0	-1.249	163
10.5	-1.184	165
11.0	-1.120	167
11.5	-1.060	169
12.0	-1.000	170
12.5	-0.942	172
13.0	-0.885	174
13.5	-0.831	176
14.0	-0.777	178
14.5	-0.725	179
15.0	-0.672	181
15.5	-0.622	183
16.0	-0.572	184
16.5	-0.524	186
17.0	-0.475	188
17.5	-0.428	190

RS	Ability	SS
18.0	-0.381	191
18.5	-0.335	193
19.0	-0.289	194
19.5	-0.244	196
20.0	-0.199	198
20.5	-0.155	200
21.0	-0.111	201
21.5	-0.068	203
22.0	-0.024	204
22.5	0.019	206
23.0	0.062	208
23.5	0.105	209
24.0	0.147	211
24.5	0.190	213
25.0	0.232	214
25.5	0.274	216
26.0	0.316	217
26.5	0.359	219
27.0	0.401	220
27.5	0.443	221
28.0	0.485	223
28.5	0.528	224
29.0	0.571	226
29.5	0.614	228
30.0	0.657	229
30.5	0.701	231
31.0	0.744	232
31.5	0.788	233
32.0	0.832	235
32.5	0.877	237
33.0	0.923	238
33.5	0.969	239
34.0	1.015	241
34.5	1.063	242
35.0	1.110	244
35.5	1.159	245

RS	Ability	SS
36.0	1.208	246
36.5	1.258	248
37.0	1.309	250
37.5	1.362	251
38.0	1.415	252
38.5	1.470	253
39.0	1.525	255
39.5	1.583	256
40.0	1.641	257
40.5	1.702	259
41.0	1.764	260
41.5	1.829	262
42.0	1.894	263
42.5	1.964	264
43.0	2.035	266
43.5	2.111	267
44.0	2.187	268
44.5	2.270	269
45.0	2.353	270
45.5	2.445	272
46.0	2.537	273
46.5	2.641	274
47.0	2.744	275
47.5	2.862	277
48.0	2.981	278
48.5	3.121	279
49.0	3.261	280
49.5	3.434	281
50.0	3.606	282
50.5	3.837	283
51.0	4.068	284
51.5	4.439	285
52.0	4.811	288
52.5	5.168	291
53.0	5.525	294

## 2004 GEPA LAL Breach Raw Score to Scale Score

RS	Ability	SS
0.0	-4.9750	104
0.5	-4.6375	106
1.0	-4.3000	107
1.5	-3.9665	109
2.0	-3.6330	112
2.5	-3.4390	114
3.0	-3.2450	116
3.5	-3.1055	118
4.0	-2.9660	120
4.5	-2.8545	122
5.0	-2.7430	124
5.5	-2.6475	126
6.0	-2.5520	128
6.5	-2.4660	130
7.0	-2.3800	132
7.5	-2.2995	134
8.0	-2.2190	137
8.5	-2.1425	139
9.0	-2.0660	141
9.5	-1.9920	143
10.0	-1.9180	145
10.5	-1.8465	147
11.0	-1.7750	148
11.5	-1.7050	150
12.0	-1.6350	152
12.5	-1.5675	154
13.0	-1.5000	155
13.5	-1.4340	157
14.0	-1.3680	159
14.5	-1.3035	160
15.0	-1.2390	162
15.5	-1.1765	163
16.0	-1.1140	165
16.5	-1.0530	166
17.0	-0.9920	168
17.5	-0.9325	169
18.0	-0.8730	171
18.5	-0.8145	172
19.0	-0.7560	173
19.5	-0.6985	175
20.0	-0.6410	176
20.5	-0.5840	178

RS	Ability	SS
21.0	-0.5270	179
21.5	-0.4705	181
22.0	-0.4140	182
22.5	-0.3575	184
23.0	-0.3010	185
23.5	-0.2450	187
24.0	-0.1890	188
24.5	-0.1325	189
25.0	-0.0760	191
25.5	-0.0195	192
26.0	0.0370	194
26.5	0.0940	195
27.0	0.1510	197
27.5	0.2085	198
28.0	0.2660	200
28.5	0.3240	201
29.0	0.3820	203
29.5	0.4410	204
30.0	0.5000	206
30.5	0.5600	207
31.0	0.6200	209
31.5	0.6805	210
32.0	0.7410	212
32.5	0.8040	213
33.0	0.8670	215
33.5	0.9305	216
34.0	0.9940	218
34.5	1.0595	219
35.0	1.1250	221
35.5	1.1920	222
36.0	1.2590	224
36.5	1.3290	225
37.0	1.3990	227
37.5	1.4710	228
38.0	1.5430	230
38.5	1.6180	231
39.0	1.6930	232
39.5	1.7710	234
40.0	1.8490	235
40.5	1.9310	237
41.0	2.0130	238
41.5	2.0985	240

RS	Ability	SS
42.0	2.1840	241
42.5	2.2735	243
43.0	2.3630	244
43.5	2.4570	245
44.0	2.5510	247
44.5	2.6480	248
45.0	2.7450	250
45.5	2.8455	251
46.0	2.9460	253
46.5	3.0485	255
47.0	3.1510	256
47.5	3.2540	258
48.0	3.3570	259
48.5	3.4635	261
49.0	3.5700	263
49.5	3.6760	264
50.0	3.7820	266
50.5	3.8900	268
51.0	3.9980	269
51.5	4.1090	271
52.0	4.2200	273
52.5	4.3365	274
53.0	4.4530	276
53.5	4.5785	278
54.0	4.7040	280
54.5	4.8420	282
55.0	4.9800	283
55.5	5.1360	285
56.0	5.2920	287
56.5	5.4750	289
57.0	5.6580	291
57.5	5.8805	293
58.0	6.1030	295
58.5	6.3915	297
59.0	6.6800	298
59.5	7.0760	300
60.0	7.4720	300
60.5	8.0505	300
61.0	8.6290	300
61.5	9.0905	300
62.0	9.5520	300



# ***Appendix D***

## ***Scale Scores with Frequencies***

## 2004 GEPA Language Arts Literacy Scale Scores with Frequencies

Scale Score	Number of Students	Percent of Students	Cumulative Number of Students	Cumulative Percent of Students	Scale Score	Number of Students	Percent of Students	Cumulative Number of Students	Cumulative Percent of Students
106	1	0.0	1	0.0	175	712	0.7	12,915	11.9
110	15	0.0	16	0.0	176	774	0.7	13,689	12.6
112	1	0.0	17	0.0	178	769	0.7	14,458	13.3
114	21	0.0	38	0.0	179	814	0.8	15,272	14.1
115	4	0.0	42	0.0	180	858	0.8	16,130	14.9
117	59	0.1	101	0.1	182	893	0.8	17,023	15.7
119	17	0.0	118	0.1	183	1,007	0.9	18,030	16.6
121	105	0.1	223	0.2	185	970	0.9	19,000	17.5
122	33	0.0	256	0.2	186	1,042	1.0	20,042	18.5
124	118	0.1	374	0.3	187	1,114	1.0	21,156	19.5
126	48	0.0	422	0.4	189	1,144	1.1	22,300	20.6
127	141	0.1	563	0.5	190	1,226	1.1	23,526	21.7
129	65	0.1	628	0.6	192	1,310	1.2	24,836	22.9
131	175	0.2	803	0.7	193	1,331	1.2	26,167	24.1
133	90	0.1	893	0.8	195	1,383	1.3	27,550	25.4
135	280	0.3	1,173	1.1	196	1,532	1.4	29,082	26.8
136	142	0.1	1,315	1.2	198	1,559	1.4	30,641	28.3
138	242	0.2	1,557	1.4	200	1,611	1.5	32,252	29.7
140	201	0.2	1,758	1.6	201	1,777	1.6	34,029	31.4
142	322	0.3	2,080	1.9	202	1,818	1.7	35,847	33.1
143	254	0.2	2,334	2.2	204	1,839	1.7	37,686	34.8
145	362	0.3	2,696	2.5	206	2,095	1.9	39,781	36.7
147	259	0.2	2,955	2.7	207	2,134	2.0	41,915	38.7
148	350	0.3	3,305	3.0	209	2,169	2.0	44,084	40.7
150	274	0.3	3,579	3.3	211	2,309	2.1	46,393	42.8
152	394	0.4	3,973	3.7	212	2,376	2.2	48,769	45.0
153	302	0.3	4,275	3.9	214	2,544	2.3	51,313	47.3
154	408	0.4	4,683	4.3	215	2,617	2.4	53,930	49.7
156	425	0.4	5,108	4.7	217	2,665	2.5	56,595	52.2
157	407	0.4	5,515	5.1	219	2,804	2.6	59,399	54.8
159	439	0.4	5,954	5.5	221	2,835	2.6	62,234	57.4
160	538	0.5	6,492	6.0	222	2,936	2.7	65,170	60.1
161	448	0.4	6,940	6.4	224	2,886	2.7	68,056	62.8
163	526	0.5	7,466	6.9	226	2,991	2.8	71,047	65.5
164	476	0.4	7,942	7.3	227	2,949	2.7	73,996	68.2
165	526	0.5	8,468	7.8	229	2,963	2.7	76,959	71.0
167	519	0.5	8,987	8.3	231	2,829	2.6	79,788	73.6
168	568	0.5	9,555	8.8	232	2,851	2.6	82,639	76.2
169	591	0.5	10,146	9.4	234	2,549	2.4	85,188	78.6
171	646	0.6	10,792	10.0	236	2,646	2.4	87,834	81.0
172	661	0.6	11,453	10.6	237	2,351	2.2	90,185	83.2
174	750	0.7	12,203	11.3	239	2,336	2.2	92,521	85.3

### 2004 GEPA Language Arts Literacy Scale Scores with Frequencies (continued)

Scale Score	Number of Students	Percent of Students	Cumulative Number of Students	Cumulative Percent of Students	Scale Score	Number of Students	Percent of Students	Cumulative Number of Students	Cumulative Percent of Students
240	1	0.0	92,522	85.3	264	213	0.2	107,710	99.3
241	2,056	1.9	94,578	87.2	265	165	0.2	107,875	99.5
242	1,902	1.8	96,480	89.0	267	128	0.1	108,003	99.6
244	1,720	1.6	98,200	90.6	269	124	0.1	108,127	99.7
245	1,636	1.5	99,836	92.1	271	73	0.1	108,200	99.8
247	1,426	1.3	101,262	93.4	272	53	0.0	108,253	99.8
249	1,231	1.1	102,493	94.5	274	53	0.0	108,306	99.9
250	1,039	1.0	103,532	95.5	276	31	0.0	108,337	99.9
252	908	0.8	104,440	96.3	278	30	0.0	108,367	99.9
254	768	0.7	105,208	97.0	279	28	0.0	108,395	100.0
255	682	0.6	105,890	97.7	281	12	0.0	108,407	100.0
257	539	0.5	106,429	98.2	283	5	0.0	108,412	100.0
258	1	0.0	106,430	98.2	285	4	0.0	108,416	100.0
259	445	0.4	106,875	98.6	287	4	0.0	108,420	100.0
260	350	0.3	107,225	98.9	289	4	0.0	108,424	100.0
262	272	0.3	107,497	99.1	290	3	0.0	108,427	100.0

**N-COUNT = 108,427 MEAN = 211.8581 STANDARD DEVIATION = 28.2136 SEM = 12.285**

## 2004 GEPA Mathematics Scale Scores with Frequencies

Scale Score	Number of Students	Percent of Students	Cumulative Number of Students	Cumulative Percent of Students	Scale Score	Number of Students	Percent of Students	Cumulative Number of Students	Cumulative Percent of Students
137	1	0.0	1	0.0	194	1,709	1.6	37,689	34.6
139	8	0.0	9	0.0	195	1,177	1.1	38,866	35.7
141	26	0.0	35	0.0	197	1,698	1.6	40,564	37.2
142	4	0.0	39	0.0	198	1,153	1.1	41,717	38.3
143	100	0.1	139	0.1	200	1,866	1.7	43,583	40.0
144	14	0.0	153	0.1	201	1,284	1.2	44,867	41.2
146	225	0.2	378	0.3	203	1,774	1.6	46,641	42.8
147	34	0.0	412	0.4	204	1,283	1.2	47,924	44.0
148	436	0.4	848	0.8	206	1,830	1.7	49,754	45.7
149	89	0.1	937	0.9	208	1,296	1.2	51,050	46.8
150	655	0.6	1,592	1.5	209	1,894	1.7	52,944	48.6
152	139	0.1	1,731	1.6	211	1,350	1.2	54,294	49.8
153	969	0.9	2,700	2.5	212	1,888	1.7	56,182	51.6
154	268	0.2	2,968	2.7	214	1,310	1.2	57,492	52.8
155	1,221	1.1	4,189	3.8	216	1,858	1.7	59,350	54.5
156	308	0.3	4,497	4.1	218	1,281	1.2	60,631	55.6
158	1,408	1.3	5,905	5.4	219	1,863	1.7	62,494	57.4
159	426	0.4	6,331	5.8	221	1,384	1.3	63,878	58.6
160	1,439	1.3	7,770	7.1	223	1,945	1.8	65,823	60.4
162	538	0.5	8,308	7.6	224	1,345	1.2	67,168	61.6
163	1,536	1.4	9,844	9.0	226	1,972	1.8	69,140	63.5
164	593	0.5	10,437	9.6	228	1,422	1.3	70,562	64.8
166	1,625	1.5	12,062	11.1	230	1,955	1.8	72,517	66.6
167	671	0.6	12,733	11.7	232	1,385	1.3	73,902	67.8
168	1,577	1.4	14,310	13.1	234	1,971	1.8	75,873	69.6
169	740	0.7	15,050	13.8	235	1,347	1.2	77,220	70.9
171	1,657	1.5	16,707	15.3	237	1,980	1.8	79,200	72.7
172	821	0.8	17,528	16.1	239	1,367	1.3	80,567	73.9
174	1,569	1.4	19,097	17.5	241	1,854	1.7	82,421	75.6
175	898	0.8	19,995	18.3	242	1	0.0	82,422	75.6
176	1,631	1.5	21,626	19.8	243	1,316	1.2	83,738	76.8
178	950	0.9	22,576	20.7	244	1	0.0	83,739	76.8
179	1,568	1.4	24,144	22.2	245	2,042	1.9	85,781	78.7
180	961	0.9	25,105	23.0	246	1	0.0	85,782	78.7
182	1,670	1.5	26,775	24.6	247	1,355	1.2	87,137	80.0
183	971	0.9	27,746	25.5	250	1,869	1.7	89,006	81.7
185	1,741	1.6	29,487	27.1	251	1,239	1.1	90,245	82.8
186	1,023	0.9	30,510	28.0	252	1,909	1.8	92,154	84.6
188	1,564	1.4	32,074	29.4	254	1,160	1.1	93,314	85.6
189	1,103	1.0	33,177	30.4	256	1,882	1.7	95,196	87.4
191	1,673	1.5	34,850	32.0	258	1,138	1.0	96,334	88.4
192	1,130	1.0	35,980	33.0	260	1,764	1.6	98,098	90.0

### 2004 GEPA Mathematics Scale Scores with Frequencies (continued)

Scale Score	Number of Students	Percent of Students	Cumulative Number of Students	Cumulative Percent of Students
262	926	0.8	99,024	90.9
263	1	0.0	99,025	90.9
264	1,719	1.6	100,744	92.5
266	887	0.8	101,631	93.3
267	1,632	1.5	103,263	94.8
269	724	0.7	103,987	95.4
271	1,267	1.2	105,254	96.6
273	574	0.5	105,828	97.1
275	1,086	1.0	106,914	98.1
276	371	0.3	107,285	98.5
278	798	0.7	108,083	99.2
280	207	0.2	108,290	99.4
281	439	0.4	108,729	99.8
282	78	0.1	108,807	99.9
284	158	0.1	108,965	100.0

**N-COUNT = 108,965 MEAN = 212.5603 STANDARD DEVIATION = 35.0312 SEM = 12.795**

## 2004 GEPA Science Scale Scores with Frequencies

Scale Score	Number of Students	Percent of Students	Cumulative Number of Students	Cumulative Percent of Students	Scale Score	Number of Students	Percent of Students	Cumulative Number of Students	Cumulative Percent of Students
133	8	0.0	8	0.0	196	2	0.0	24,391	22.4
137	9	0.0	17	0.0	197	1,989	1.8	26,380	24.2
139	1	0.0	18	0.0	198	1,122	1.0	27,502	25.3
141	16	0.0	34	0.0	200	2,068	1.9	29,570	27.2
143	2	0.0	36	0.0	202	1,172	1.1	30,742	28.2
145	44	0.0	80	0.1	203	2	0.0	30,744	28.2
146	8	0.0	88	0.1	204	2,046	1.9	32,790	30.1
148	75	0.1	163	0.1	205	1,334	1.2	34,124	31.4
150	14	0.0	177	0.2	206	1	0.0	34,125	31.4
152	198	0.2	375	0.3	207	2,075	1.9	36,200	33.3
154	28	0.0	403	0.4	208	1,281	1.2	37,481	34.4
156	336	0.3	739	0.7	209	1	0.0	37,482	34.4
158	46	0.0	785	0.7	210	2,121	1.9	39,603	36.4
159	527	0.5	1,312	1.2	211	1,360	1.2	40,963	37.6
160	1	0.0	1,313	1.2	213	2,075	1.9	43,038	39.5
161	94	0.1	1,407	1.3	214	1,403	1.3	44,441	40.8
163	768	0.7	2,175	2.0	216	2,073	1.9	46,514	42.7
165	161	0.1	2,336	2.1	217	1,432	1.3	47,946	44.1
166	1,063	1.0	3,399	3.1	219	2,027	1.9	49,973	45.9
167	5	0.0	3,404	3.1	220	1,440	1.3	51,413	47.2
168	228	0.2	3,632	3.3	221	1	0.0	51,414	47.2
170	1,313	1.2	4,945	4.5	222	2,058	1.9	53,472	49.1
172	338	0.3	5,283	4.9	223	1,490	1.4	54,962	50.5
173	1,622	1.5	6,905	6.3	225	2,121	1.9	57,083	52.4
174	4	0.0	6,909	6.3	227	1,537	1.4	58,620	53.9
175	425	0.4	7,334	6.7	228	2,077	1.9	60,697	55.8
176	4	0.0	7,338	6.7	229	2	0.0	60,699	55.8
177	1,787	1.6	9,125	8.4	230	1,506	1.4	62,205	57.2
178	4	0.0	9,129	8.4	231	1,950	1.8	64,155	58.9
179	550	0.5	9,679	8.9	232	1,535	1.4	65,690	60.4
180	1,951	1.8	11,630	10.7	233	1	0.0	65,691	60.4
181	6	0.0	11,636	10.7	234	1,997	1.8	67,688	62.2
182	707	0.6	12,343	11.3	235	1,484	1.4	69,172	63.6
184	2,089	1.9	14,432	13.3	237	2,036	1.9	71,208	65.4
185	825	0.8	15,257	14.0	238	1,421	1.3	72,629	66.7
186	5	0.0	15,262	14.0	239	1,888	1.7	74,517	68.5
188	2,068	1.9	17,330	15.9	241	1,430	1.3	75,947	69.8
189	875	0.8	18,205	16.7	242	1,984	1.8	77,931	71.6
191	2,051	1.9	20,256	18.6	244	1,440	1.3	79,371	72.9
192	968	0.9	21,224	19.5	245	1,881	1.7	81,252	74.7
194	2,093	1.9	23,317	21.4	246	1,407	1.3	82,659	75.9
195	1,072	1.0	24,389	22.4	248	1,935	1.8	84,594	77.7

### 2004 GEPA Science Scale Scores with Frequencies (continued)

Scale Score	Number of Students	Percent of Students	Cumulative Number of Students	Cumulative Percent of Students
249	1,264	1.2	85,858	78.9
250	1,689	1.6	87,547	80.4
251	2	0.0	87,549	80.4
252	1,283	1.2	88,832	81.6
253	1,626	1.5	90,458	83.1
254	1,167	1.1	91,625	84.2
256	1,558	1.4	93,183	85.6
257	1,107	1.0	94,290	86.6
258	1,400	1.3	95,690	87.9
259	1,045	1.0	96,735	88.9
261	1,300	1.2	98,035	90.1
263	1,006	0.9	99,041	91.0
264	1,239	1.1	100,280	92.1
265	850	0.8	101,130	92.9
266	1,104	1.0	102,234	93.9
267	726	0.7	102,960	94.6
269	987	0.9	103,947	95.5
270	684	0.6	104,631	96.1
271	789	0.7	105,420	96.9
272	531	0.5	105,951	97.3
273	585	0.5	106,536	97.9
274	420	0.4	106,956	98.3
276	486	0.4	107,442	98.7
277	332	0.3	107,774	99.0
278	305	0.3	108,079	99.3
279	187	0.2	108,266	99.5
280	222	0.2	108,488	99.7
281	108	0.1	108,596	99.8
282	116	0.1	108,712	99.9
283	47	0.0	108,759	99.9
284	41	0.0	108,800	100.0
286	19	0.0	108,819	100.0
288	15	0.0	108,834	100.0
291	3	0.0	108,837	100.0
294	4	0.0	108,841	100.0

**N-COUNT = 108,841 MEAN = 222.1660 STANDARD DEVIATION = 29.9309 SEM = 10.565**

# ***Appendix E***

## ***Limited English Proficiency (LEP) and Special Education (SE)***

**LIMITED ENGLISH PROFICIENT (LEP)**

A limited English proficient student is a student whose native language is one other than English. This student has sufficient difficulty speaking, reading, writing, or understanding the English language, as measured by an English language proficiency test, so as to be denied the opportunity to learn successfully in the classroom where the language of instruction is English. The codes for LEP are:

- < = LEP student entered a language assistance program AFTER July 1, 2003, and is currently enrolled in the program.
- 1 = LEP student entered a language assistance program BETWEEN July 1, 2002, and June 30, 2003, and is currently enrolled in the program.
- 2 = LEP student entered a language assistance program BETWEEN July 1, 2001, and June 30, 2002, and is currently enrolled in the program.
- 3 = LEP student entered a language assistance program BEFORE July 1, 2001, and is currently enrolled in the program.

**SPECIAL EDUCATION (SE)**

There are 14 codes for Special Education categories. The categories are:

- A. Auditorily Impaired
- B. Other Health Impaired
- C. Communication Impaired
- D. Emotionally Disturbed
- E. Cognitively Impaired
- F. Multiply Disabled
- G. Traumatic Brain Injury
- H. Orthopedically Impaired
- I. Specific Learning Disability
- J. Social Maladjustment
- K. Visually Impaired
- L. Speech-Language Services Only
- M. Autistic
- N. Refers to one of the above. This is a default code when an IEP student failed to provide the specific information listed above.

# ***References***

**REFERENCES**

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education (AERA, APA, NCME) (1999) *Standards for Educational and Psychological Testing*. Washington, DC: Author.

American Psychological Association, American Educational Research Association, & National Council on Measurement in Education (1974) *Standards for Educational and Psychological Testing*. Washington, DC: American Psychological Association.

Camilli, G. & Shepard, L.A. (1994). *Methods for Identifying Biased Test Items*. Thousand Oaks, CA: Sage.

Cizek, G.J. (1995). *Standard setting as Psychometric Due Process*. Paper presented at the Annual Meeting of the National Council on Measurement in Education, San Francisco.

*Cycle II Criterion-Based Holistic Scoring: A Writing Handbook*. Developed by the New Jersey Department of Education, September 2004.

DeMauro, G.E. (1995). *Construct Validation of Minimum Competence*. Paper presented at the Annual Meeting of the National Council on Measurement in Education, San Francisco.

*Directory of Test Specifications and Sample Items for the Elementary School Proficiency Assessment (ESPA), Grade Eight Proficiency Assessment (GEPA), and High School Proficiency Assessment (HSPA) in Language Arts Literacy*. Developed by the New Jersey Department of Education, February 1998.

*Directory of Test Specifications and Sample Items for the Grade Eight Proficiency Assessment (GEPA) and the High School Proficiency Assessment (HSPA) in Mathematics*. Developed by the New Jersey Department of Education, February 1998.

*Directory of Test Specifications and Sample Items for the Grade Eight Proficiency Assessment (GEPA) and High School Proficiency Assessment (HSPA) in Science*. Developed by the New Jersey Department of Education, February 1998.

*Grade Eight Proficiency Assessment (GEPA) Standard Setting Study Report of Activities. 1999 Language Arts Literacy and Mathematics, and 2000 Science*. Developed for the New Jersey Department of Education by National Computer Systems, November 2000.

Holland, P.W. & Thayer, D.T. (1986). *Differential Item Functioning and the Mantel-Haenszel Procedure* (Technical Rep. No. 86-69). Princeton, NJ: Educational Testing Service.

Jaeger, R.M. & Mills, C.N. (1998). An integrated judgment procedure for setting standards on complex large-scale assessments. Paper presented at that annual meeting of the American Educational Research Association, San Diego, CA.

- Kolen, M.J. & Brennan, R.L. (1995). *Test equating*. New York: Springer-Verlag.
- Livingston, S.A. & Zieky, M.J. (1982) *Passing Scores*. Princeton, NJ: Educational Testing Service.
- Masters, G. N & Wright, B. (1997). The partial credit model. In W. J. van der Linden & R.K. Hambleton, (Eds.), *Handbook of Modern Item Response Theory*. New York. Springer-Verlag.
- Mehrens, W.A. & Lehmann, I.J. (1991) *Measurement and Evaluation in Education and Psychology* (4th ed.). New York: Holt, Rinehart and Winston.
- Messick, S. (1980). Test validity and the ethics of assessment. *American Psychologist*, 35, 1012 – 1027.
- Millman, J. & Greene, J. (1989) The specification and development of tests of achievement and ability. In R. L. Linn (Ed.), *Educational Measurement* (3rd ed., pp. 335 – 366) New York: American Council on Education.
- SAS Institute (1985). *User's Guide: Statistics* (vol. I and II). Cary, NC: Author.
- School and District Guidelines: Interpretation and Use of GEPA Results*. Developed by the New Jersey Department of Education, June 2004.
- Van Der Linden, W.J. & Hambleton, R.K. (Eds.)(1997). *Handbook of Modern Item Response Theory*. New York: Springer-Verlag.
- Webb, M.W. & Miller, E.R. (1995). *Standard Setting on Constructed-Response Items*. Paper presented at the Annual Meeting of the National Council on Measurement in Education, San Francisco.
- Zieky, M. (1993) Practical questions in the use of DIF statistics in test development. In P.W. Holland and H. Wainer (Eds.), *Differential Item Functioning*. Hillsdale, NJ: Lawrence Erlbaum.





Copyright © 2006 New Jersey Department of Education  
All rights reserved.

